

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

v1 protein - protein search, using sw model

run on: January 30, 2004, 13:12:38 ; Search time 22 Seconds
(without alignments)
2313.635 Million cell updates/sec

title: US-09-769-699-2

erfect score: 6294
sequence: 1 MENTQKTVTPTGLGVYA.....DELFLSGIPKIGNITMEM 1203

oring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 328717 seqs, 42310858 residues

otal number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/1/aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/1/aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/1/aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/1/aa/PCTUS.COMB.pep:*
- 6: /cgn2_6/ptodata/1/aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	3054	48.5	1452	4	US-09-127-227-2
2	183	2.9	35	2	US-08-607-332B-9
3	183	2.9	35	3	US-09-338-876-9
4	131	2.1	757	3	US-09-413-814-84
5	123	2.0	1074	4	US-09-071-035-358
6	123	2.0	1074	4	US-09-071-035-394
7	118.5	1.9	370	4	US-09-252-991A-16913
8	111	1.8	5215	3	US-07-542-734C-4
9	106.5	1.7	3567	2	US-08-439-009A-4
10	106.5	1.7	3567	3	US-07-731-157A-7
11	106	1.7	774	4	US-08-541-780-7
12	106	1.7	774	2	US-09-107-532A-3855
13	105.5	1.7	1095	4	US-08-633-760-46
14	105	1.7	774	1	US-08-633-760-48
15	105	1.7	774	4	US-08-931-608A-5
16	104.5	1.7	635	4	US-08-019-870-5
17	104.5	1.7	774	1	US-08-019-870-3
18	103	1.6	774	1	US-08-633-760-44
19	103	1.6	729	1	US-08-070-165F-6
20	102	1.6	729	2	US-08-885-418-6
21	102	1.6	774	1	US-07-747-901A-3
22	102	1.6	774	1	US-07-935-312-3
23	102	1.6	774	1	US-08-633-760-50
24	102	1.6	774	1	US-08-019-870-1
25	101.5	1.6	773	1	US-08-019-870-6
26	101.5	1.6	773	1	US-08-019-870-8
27	101.5	1.6	774	1	US-08-019-870-8

28	101.5	1.6	774	1	US-08-019-870-11	Sequence 11, Appl
29	101.5	1.6	774	1	US-08-633-760-52	Sequence 52, Appl
30	101	1.6	995	4	US-09-687-931A-1	Sequence 1, Appl
31	101	1.6	1024	4	US-09-582-737-48	Sequence 48, Appl
32	101	1.6	1466	4	US-09-282-537-20	Sequence 20, Appl
33	101	1.6	7257	3	US-09-335-409-5	Sequence 5, Appl
34	101	1.6	7257	4	US-09-568-102-5	Sequence 5, Appl
35	101	1.6	7257	4	US-09-567-969-5	Sequence 5, Appl
36	101	1.6	7257	4	US-09-568-480-5	Sequence 5, Appl
37	101	1.6	7257	4	US-09-568-486-5	Sequence 5, Appl
38	101	1.6	7257	4	US-09-568-472-5	Sequence 5, Appl
39	101	1.6	7257	4	US-09-567-899-5	Sequence 5, Appl
40	100.5	1.6	1321	1	US-08-261-822A-3	Sequence 3, Appl
41	100.5	1.6	1321	5	PCT-US95-07744A-3	Sequence 3, Appl
42	100.5	1.6	2616	6	5206163-3	Patent No. 5206163
43	99.5	1.6	600	4	US-09-252-991A-29817	Sequence 29817, A
44	99	1.6	915	1	US-08-328-322-5	Sequence 5, Appl
45	98.5	1.6	801	4	US-09-351-150A-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-09-127-227-2
; Sequence 2, Application US/09127227
; Patent No. 6399354
; GENERAL INFORMATION:
; APPLICANT: David M. Knipe
; APPLICANT: Travis J. Taylor
; APPLICANT: Elizabeth McNamee
; TITLE OF INVENTION: Replication-Competent Virus Expressing A
; FILE REFERENCE: HU98-05
; CURRENT APPLICATION NUMBER: US/09/127,227
; CURRENT FILING DATE: 1998-07-31
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1452
; TYPE: PRT
; ORGANISM: herpesvirus
US-09-127-227-2

Query Match	48.5%;	Score	3054;	DB 4;	Length	1452;	
Best Local Similarity	49.8%;	Pred. No.	2.7e-304;	Indels	48;	Gaps	17;
Matches	609;	Conservative	200;	Mismatches	366;		
1	MENTQKTVT---VPTGPLGVY--ACRVEDLDLEISFLAARSTDS	LALLPLMRNLTV	55				
1	METPKTATIKVPPGPLVYARACPSGIEL--LALLSARSGDADV	AVAPLVVGLTVE	58				
56	KFTTSSIANVSGARTTGLAGAGITLKTTSHFYSPVVFHGGKHL	VPSAAPNLTRACNA	115				
59	SGFEANVAVVGSRTTGLGTAVSLKLTTPSHYSSVVFHGGKHL	DPSTQAPNLTRLCER	118				
116	ARERFGSRCCGPPVDCGAVETTGAEICTRLGLEPENTILYLV	VLTALFKAEVFCNVLHY	175				
119	ARRHFGESDYTPRGDLKHETTGALCERLGLDPRALLYLV	VTEGFKAVCINNTLHL	178				
176	CGLDIVHINHGDIIRIFLPVOLFPMDVNRILVPPFNTHRS	IGSGFYPTPEVNTGLCH	235				
179	CGSKDVTIGGAEVHRIPVYPLQLEFMPDFSRVIAEPFNANHS	IGENTYLPFPFNRLNR	238				
236	LHDCVIAFMAVALRVNVTAVAGAAHLAFDENHGAALPPDIT	YTYFOSSSGTTTAR	299				
239	LLFEAVGPAVALRCRNDVAVARAAHLAFDENHGAALPADIT	FATFASQG--KTPR	299				
296	GARRNDVNSTKSPSGGFERELASIMAADTALHAEVIFNTGI	YEEPTDIIKENPMFIGM	351				
297	GGR-----DGGKGAGGFGFQRLASVWAGDAALAESIVMA	VDEPPTDLSAMPLECGQ	351				
356	EGTLPLNALGSYTAEVAGVICAMVFPNSALYLTEVEDSGMT	EAKDGGPGSFNRFYQF	415				

09/769,699
2.21.04
SEARCH NOTES

352 DTAARANAVGAYLARAAGLVGVFNSALHTEVDAGPADPKDHSK-PSFYRFFLV 410
416 AGPHLANPQTRDCHVL-----SQSTGSSNTEFSDYDYLALICGFGAPILARLFLYL 468
411 PGTHVAANPQVDRGHVVFEGFRTAPLVGGTQ-ETAGEHLAMLCGFSALLAKMLFYL 469
469 ERCDAGFTGGHG-DALXYVTGTSDSEIPSCLEKTRPVCAHTVHRLRQRPRFQCAT 527
470 ERCDGVTGQEMDVFRYVADSNQTDVPCNLCTFDTRHACVHTLRLARHPKPSAA 529
528 RQPIGVFTMNSQVSDCDPLGNYPYLILKPGQOCTEAAKATMQDTYRATLEFLIDLEQ 587
530 RGAIGVFTMNSMYSDDCVLGNVYAFSAKLR-ADGSETARTIMQETTYRAATERVMAELET 588
588 BRLLDRGAPCSSEGLSSVIVDHPTRFRLLDRLARIEQTTTFQFMKVLVETRDYKIREGLS 647
589 LOYVDQAVPTANGRLTITITREALHTVNNVQVQVREVEQLMNLVEGKFRDGLG 648
648 EATHSMALTFDYPGAFCEITNLFVRLHAWQDLALSOCHGVYQCGVEGRNFRNQFQ 707
649 EANHAMSUTLDYACPCPLQLLGRRLNLAVIDLALSOCHGVYQCGVEGRNFRNQFQ 708
708 PVLRRFVDLNGGFISTRSTVTLSEG-PVSAENPTLGDAPAGRTFDGDLARVSVEVI 766
709 PVLRRFVDLNGGFISTRSTVTLSEG-PVSAENPTLGDAPAGRTFDGDLARVSVEVI 766
767 RDIVKRVNVSNGNTNLSEARARLVGLASAYORQEXRVDMLHGLGFLKQPHGLLFF 826
769 KELRYKSVRLFAGANASAAKARVASLQSAQYKPDKXVDILLGPLGLKQPHAAIFP 828
827 RQMPNSKSPNQFWTLLORNOPADKLTHEETITIAVKRFTVEEYAAINFILPTCI 886
829 NGKPPGSSQNPQFWTALORNLPARLSREDIETIAFIKKFLSDYGAINFILAPNV 889
887 GELAFYMANLTKYCDHSOYLINTLSIIGARRPRDPSSVLEHWIRKDVTSAADTETQA 946
889 SELAMYMANQILRYCDHASTFYINTLTAIAGSRPPSPVQAAA--SAOGGAGLEAGA 945
947 KALLEKTENLPELWTAFTSTHVLRAANQPMVVLG-SISKYHGAAGNNRVFQAGHWSG 1006
946 RALMDAVIDAHPGAWTSMFASCNLLRPVNAARPMVVLG-SISKYGMAGNDRVFOAGNWS 1005
1007 LINGKXVCLFTRTRFLIACPGGTCPTVGPSSGNRETTLSDOVRGIIVSGAMVQ 1066
1006 LMGGKXACPLLIIDTRKFLVACPRAGFVCAASNLGGGAHSSCEQURGIISGGAVA 1065
1067 LAIYATVVRVAVGARQHMVAFDDWLSITDDEFLARDLELHDQIIQTLETPMTVEGAL--- 1123
1066 SSVFVATVKSGLPRTQOLQIEDWALLEDEYLSSEMMELTARALRGNGENSTDAALEVA 1125
1124 -EAVKILDEKTAGDGETTNLAFNFD--SCFPHDTSNVLNIGSNISSGTVPGKRP 1180
1126 HEAEALVSQLNAGE-----VFNFGDGC-----DDNATPFGGPGAGPAPAGAKRA 1173
1181 PEDDELFDLSGIPKHNITMEM 1203
1174 FHGDDPFG-EGPDKKGDJLDM 1195

RESULT 2
US-08-807-332B-9
; Sequence 9, Application US/08807332B
; Patent No. 5959074
; GENERAL INFORMATION:
; APPLICANT: Dreyfus, David H.
; APPLICANT: Gelfand, Erwin W.
; TITLE OF INVENTION: PRODUCTS AND PROCESSES FOR REGULATION OF
; TITLE OF INVENTION: GENE RECOMBINATION
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross
; STREET: 1700 Lincoln St., Suite 3500

CITY: Denver
STATE: CO
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/807,332B
FILING DATE: 28-FEB-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kovarik, Joseph E.
REGISTRATION NUMBER: 33,005
REFERENCE/DOCKET NUMBER: 2879-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/863-9700
TELEFAX: 303/863-0223
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 35 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-807-332B-9
Query Match 2.9%; Score 183; DB 2; Length 35;
Best Local Similarity 100.0%; Pred. No. 1.4e-11;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 545 DPLGNVAPYLILRKPGDQTEAKATMQDTYRATLE 579
Db 1 DPLGNVAPYLILRKPGDQTEAKATMQDTYRATLE 35
RESULT 3
US-09-338-876-9
; Sequence 9, Application US/09338876
; Patent No. 6187584
; GENERAL INFORMATION:
; APPLICANT: Dreyfus, David H.
; APPLICANT: Gelfand, Erwin W.
; TITLE OF INVENTION: PRODUCTS AND PROCESSES FOR REGULATION OF
; TITLE OF INVENTION: GENE RECOMBINATION
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/338,876
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/807,332
FILING DATE: 28-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Kovarik, Joseph E.
REGISTRATION NUMBER: 33,005
REFERENCE/DOCKET NUMBER: 2879-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/863-9700

TELEFAX: 303/863-0223

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 35 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

IS-09-338-876-9

Query Match 2.9%; Score 183; DB 3; Length 35;

Best Local Similarity 100.0%; Pred. No. 1.4e-11;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

545 DPLGNVAPYLILRKPGDQTEAAKATMDTYRATLE 579

1 DPLGNVAPYLILRKPGDQTEAAKATMDTYRATLE 35

RESULT 4

JS-09-413-814-84

Sequence 84, Application US/09413814

Patent No. 6225064

GENERAL INFORMATION:

APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH

APPLICANT: Bristol-Myers Squibb, Co.

APPLICANT: Beyer, Stefan

APPLICANT: Bioecker, Helmut

APPLICANT: Brandt, Petra

APPLICANT: Cinc, Paul M

APPLICANT: Dougherty, Brian A

APPLICANT: Goldberg, Steven L

APPLICANT: Hofle, Gerhard

APPLICANT: Mueller, Joachim

APPLICANT: Reichenbach, Hans

TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or

TITLE OF INVENTION: heteropolyketide compounds

FILE REFERENCE: PCT/US 99/23535

CURRENT APPLICATION NUMBER: US/09/413,814

EARLIER FILING DATE: 1999-10-07

EARLIER APPLICATION NUMBER: DE 198 46 493.2

NUMBER OF SEQ ID NOS: 109

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 84

LENGTH: 757

TYPE: PRT

ORGANISM: Sorangium cellulosum

JS-09-413-814-84

Query Match

Best Local Similarity 21.1%; Pred. No. 0.001;

Matches 146; Conservative 84; Mismatches 251; Indels 212; Gaps 37;

QY

54 VEKFTSSLAWSGART-----TGLAGAGITLKLTTSHFYPSVFVFGGKHLVLPSSAAPNL 109

129 LERTLPRAVSIVADARAALTSAAVIVAS-----LPASAAA-- 167

QY

110 TRACNAARERFGSRGCGPPVDGAVETTGAEICTRLGLEPENTILYLVTALPKFAVPMC 169

168 -----ALQRLRWIATQPS-PGPIEGFGA-----ALRPSV----- 198

QY

170 NVFLHY-----GGLDIVHNGDVIRPLPFVQLFMPDVNRLVPDPFNTHRSIGSGFVY 224

199 -AFLOYTSGSTGEPKGVMLTHGNDL-----LHNSRLIAHGFDLTSPDPV-----GVIW 244

QY

225 PTPYNTGLC-----HLIHDVIAPIAVALR-----VENVTAV-----ARGAHLAFD-- 267

245 LPYHDMGLIGILQALYRIRVALMGPLFLQPLQPMKWLRAVSGALGASVGGGNFAYDLC 304

QY

268 -----ENEGAVLPDITVTYFQSSSGTGTARGARRNDVNSTKPSGSGGFERRLASIVA 323

305 VRKSEBERAALD-----LRSNEVAFPTGAEVPVRAOTLDRFARAPAVSGFRRE--AFYP 355

QY 324 ADTALHAEIFNTGIYEETPTDIKEWPMFIGNEGTLPRLNALGSYTRVAGVIGAMVF-- 381
DB 356 CYGLAEATLIVSGARAEAPV-----LARLAPEEVLGRAVASAAE--GARVFIG 403
QY 382 -----SPNSALYLTEVEDSGMTEAKDGGPGSFNRFYFAGPHLA-----ANPQTDQDGHV 432
DB 404 SGRALDPRA---VAIVDPAG---NELGPG-EIGELW-VSGPSVAVGYWGRPE----- 447
QY 433 LSSQSTGSSNTEFSVDYLALICGFGAPLLARLLFYLERCDAGAFTHGGHDALKYVTGT-- 490
DB 448 -----ETEATFGATLAGSAAPR-----YLRTGDLGFLRGGE-----LFVVGSRK 486
QY 491 -----FDSEIFCSLCEKH--TRPVC--AHTTVHRLRQM-----PRFGQATRFQ 529
DB 487 DLILGRNHFPDIEKTVESSHRAVRPCSRAPSVHEGEERLAVVCEVDPRVAADPRE 546
QY 530 PIGVFTGMSQYSDCDPLGNYPYLILRKPGQTEAAKATMDTYRATLERLIDLEQER 589
DB 547 IVAAREAVTAHQ-----LVAAHVALIAPGALPKTSSGKVRRECRRAFL-----DALGER 598
QY 590 -----LLDRGAPCSSEGLSSVIVDPTFRRLDTLRARIETTTQFMKVLVETRDYKI 642
DB 599 HVAFAPELLDDAGFPDD---APPETEEPSGRSLDALRS-----TLARLRLDAGQIDDAL 651
QY 643 ---REGI-SEATHSMALTFDPYSGAFCPITNFI 671
DB 652 PISRFGLDSLAAVELQHAFQVTRGRAIPLTSIL 684

RESULT 5

US-09-071-035-358

Sequence 358, Application US/09071035

Patent No. 6448043

GENERAL INFORMATION:

APPLICANT: Gil H. Choi

TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 496

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071,035

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brookes

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 358:

SEQUENCE CHARACTERISTICS:

LENGTH: 1074 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-071-035-358

Query Match 2.0%; Score 123; DB 4; Length 1074;
 Best Local Similarity 18.3%; Pred. No. 0.013;
 Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps 52;

QY 250 RVNVTAVARGAAHLAFDENHEGAVLPDDITYTFQSSSGSTTTARGARRNDVNSTKPS 309
 DB 67 RTSLIAYENGAKQTVFC-IEFGVSIPTVTHGY-----QKNPL 104

QY 310 PSGGFERRLASIM--AADTALHAENVFNTGIYEE-----TPDKEWPMFI 353
 DB 105 PMSDKAKLVSLWEKAGTDIDTNMVAQKMIWEENVGYKLHSIKRLGGASVDIK-----158

QY 354 GMEGTPLRNALGSYAR-----VAGVIG-AMVFPNSALYLTEVEDSGMTEAKDGGP 405
 DB 159 SIEGKINK--AIEEYQKPSFHTVTKILGOSTLIDKXNLNLSFDPKVVQNTA-----211

QY 406 GPSFNRFYQFAGPHLAANPQTRDGHVLSQSGNSTEFSVDYLALICGFGAPLLARLL 465
 DB 212 ---NIDYRVIGNQLVLP-----NSNSKSGTLLKKSAGTGTP-VAYKK 251

QY 466 FYLERCDAGFTGGHGDALKVYTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQMRPFQ 525
 DB 252 AGLQVWAGALDKPNTYAIAKINVTKGS-LKIKKIDKESGDIVETVPHL-----DFGK 304

QY 526 A-----TRQPIGVFGTWNQSVQDCDPLGNYA-----PYLILRKPQDTEAAKATM 570
 DB 305 ALPSKDVITDKGI-----SILDGIHGKTVITEKSVDPYMDITPMAATIKAGETI 358

QY 571 QTYRATLERIFIDLEQBRLLDRGAPCSSEGLS-----SVVDHPT---PRRLDITLRA 621
 DB 359 SMTSNMRKQGOILLEKTG-VETGDLNDNYSLAGNTFAIRKDSPAGEIVQEIITDEKG 417

QY 622 RIEQTTQFMKLVETRDYKIREGSEATHSMALTFDP-----YSGAFCEITNFLVKRT 675
 DB 418 RAE--TPKELANALELGYVTE--TKSNGFVNTFKTKVELKYANQTVALVTSNVKGQ 473

QY 676 HLAVVQDIAL-----SQCHCVFYQQV-----EGRNFRNQFQVLRERRFVLD 717
 DB 474 NQEIETGTTLKEDKDTGNESQCKAEFKGAEVLTFTAKDGOAVKWEAFK-----TEL 526

QY 718 FNGGFSIRTSITVLE-CPYSAPNPTLQ-----DAPAGRTDGDGLARVSVB-----764
 DB 527 VKGTASDETVTLADEKNQVAVXHLAINEYFWQETKAPGYTLDKTYPVSIKKVNE 586

QY 765 ----VIRDVRNRRV-----PSGNTNLSEAA-----RARLVGLASAYQROEKR 805
 DB 587 KNAVITRDVTAKEQVIRGDFDFKAGSADGTAETGENDLSFKVSPLEGTXEITGAEDKA 646

QY 806 VMLHGAIGF-----LLQPHGLLPPGM-----PPNSKS-----PNP 838
 DB 647 TTACNEQLGFDGKFNLPYGDYLLSEIEA---PEGFKITPLEIRSTFKENKDDVAKS 703

QY 839 QWFWTLLQNO-----MPADKLTHEEITIAAVKFTBEYAAININLP--PTCIGE 888
 DB 704 EYVFHTEGQKQPKWVTVPEKLTNNE-----FSVSLNRLMLYDLPEKEDSLTS 754

QY 889 LAQFYMANILKYCHRSOVLINTLSIITGARRRDPSSVLHWIRKO---VTSAAIDFTQ 945
 DB 755 LATWDGKNKLANLTDPE-LVDKL-----RYNLHEIKEDWYVVAQIDVEA- 799

QY 946 AKALLEKTENLPETWTATFTST-----HLVRAAMNORPMVVLGISISKVHGA 992
 DB 800 TQAAQEKDEKAPVIAETITLANKEKGTWKILKLTAEQ-----VLDKSVLFNYV 853

QY 993 AGNRRVFAQGNWSGLNGKNVCPLETFDTRFRFIACPRGGFICPVTFSSGNRETTLS 1052
 DB 854 YENKVAFEAGNE-----PVA-----KDALSLN 875

QY 1053 QVPGIIVSGGAMVQLAIYATVRAVGAR-----AQHWAFFDMLSLTDDDEFLARDL-- 1102
 DB 876 QAQ-----TVNCTIERHVSIOQKAHLEDGSGQTFTHGVDNDMFDVDSVTHVDLD 923

QY 1103 --BELHQIQTLETPWTVEGALEAVKILDE-----KTTAGD-----GETPTNLAFFND 1149

DB 924 GSKEAFETILVALLPDGINKIWSKGLIEHVNDEKFTKTVLAEKVDTGKYPEGTKETP- 982

QY 1150 SCEPSHDTTSNV-----LNISGNSISGTVPGKRPPEDE 1185

DB 983 -TEINYEKDGNVGNKHEDLKEKSTLTTPKEVPTIPSTPKQPE 1024

RESULT 6
 US-09-071-035-394
 ; Sequence 394, Application US/09071035
 ; Patent No. 6448043
 ; GENERAL INFORMATION:
 ; APPLICANT: Gil H. Choi
 ; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
 ; NUMBER OF SEQUENCES: 436
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/071,035
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: A. Anders Brookes
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB36992
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 394:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1074 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-071-035-394

Query Match 2.0%; Score 123; DB 4; Length 1074;
 Best Local Similarity 18.3%; Pred. No. 0.013;
 Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps 52;

QY 250 RVNVTAVARGAAHLAFDENHEGAVLPDDITYTFQSSSGSTTTARGARRNDVNSTKPS 309
 DB 67 RTSLIAYENGAKQTVFC-IEFGVSIPTVTHGY-----QKNPL 104

QY 310 PSGGFERRLASIM--AADTALHAENVFNTGIYEE-----TPDKEWPMFI 353
 DB 105 PMSDKAKLVSLWEKAGTDIDTNMVAQKMIWEENVGYKLHSIKRLGGASVDIK-----158

QY 354 GMEGTPLRNALGSYAR-----VAGVIG-AMVFPNSALYLTEVEDSGMTEAKDGGP 405
 DB 159 SIEGKINK--AIEEYQKPSFHTVTKILGOSTLIDKXNLNLSFDPKVVQNTA-----211

QY 406 GPSFNRFYQFAGPHLAANPQTRDGHVLSQSGNSTEFSVDYLALICGFGAPLLARLL 465
 DB 212 ---NIDYRVIGNQLVLP-----NSNSKSGTLLKKSAGTGTP-VAYKK 251

QY 466 FYLERCDAGFTGGHGDALKVYTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQMRPFQ 525

```

b 252 AGLQTVAGALDKPNTYAIAKINVTGKS-LKIKIDKESGDIVETVPHL-----DFGX 304
y 526 A-----TRQPIGVTGWSQVSCDPLGNYA-----PYLILKPKGQTEAAKATM 570
b 305 ALPKSDVTTDKGI-----SILDGIPIGKVTITEKSVPPPYMIDTTPMAATIKAGETI 358
y 571 QDVTATLIERIFIDLEQERLLDRGAPCSSEGLS-----SVVDHPT---PRRILDTLRA 621
b 359 SMTSKNMRQKGILLKTKG-VETGTDLWMDNYSLAGNTFAIRKDSPAGEIVQEIITDEKG 417
y 622 RLEQTTQFMKVLVETRDYKIREGLSEATHSNALTFD-----YSGAFCDITNPLVKRT 675
b 418 RAE--TPKELANALELGTYYVE--TKSGNFVNIFKTKVELKIANQTVALTNSVKGQ 473
y 676 HLAVVQDLAL-----SOCHCVFYGOQV-----ECRNFNQFQVPLRRRFVDL 717
b 474 NOEITGETTLTKEDKDTGNSGKAEPKGAETLFTAKDGOAVKWEAFK-----TEL 526
y 718 FNGGFISTRSITVTLSE-GPVSAPNPTLQ-----DAPAGTDDGLDARVSVE----- 764
b 527 VKGTASDETVTLADEKNQAVKHLAINEYFWQETKAPEGTTLDETYPVSIKKVONNE 586
y 765 ----VIRDIRVNRVV-----PSGNCNTLSEAA-----EARLVGLASAYQROEXR 805
b 587 KNAVITRDVTAKEQVIRGDFOFKFGAGSADGTAETGFDLSEKVSPLBGTXEITGAEDKA 646
y 806 VMLHGALEF-----LLKQPHGLLPFRGM-----PNSKS-----PNP 838
b 647 TTACNEQLGQDYGKFNPLPYDYLLEIEA-----PEGFQKITPLEIRSTFKENKDDYAKS 703
y 839 ONFWITLLORNO-----MPADKLTHEITIAAVKRPTEEYAAINFILP--PTCIGE 888
b 704 EVFVITTEGQKQPKQVTVPEKLTNNE-----FSVSLNELMLYDLPEKEDSLTS 754
y 889 LAQFYMANLILKYDSQVLYNTLSIITGARPRDPSSVLHWIKD---VTSAADITQ 945
b 755 LATWKDGNKNTLDFTE-LVDKL-----RYNLHEIKEDWYVVAQIDVEA- 799
y 946 AKALLEKTENLELMTTAPTST-----HLVRAAMNQRPMVLGKISKVGHA 992
b 800 TKAQEKDKAPVVAETATLANEKGTGWILHUKJTAEQ-----VLKSLVLFNVY 853
y 993 AGNNRVFQAGNSGLNGGRNVCPLETFDTRFIIACPRGGFICPVTPGSSGNRETTLS 1052
b 854 YENKVAFAEGNE-----PVA-----KDALNN 875
y 1053 QVRGIIVSGAMVQALAIYATVAVCAR-----AQHAFDDWLSLTDDEFLARDL-- 1102
b 876 QAK-----TVNCTIERHVSIOTKAHLDEGSDTFTHGVDYMDPDDVSVTHVDLD 923
y 1103 --BELHDOIOTLETPWTVEGALEAVKILDE-----KTTAGD---GETPTNLAFNFD 1149
b 924 GSKEAFETILYALLPDGTNKEIKWSKIBHEVNDKXETKVLAEKVDIGKYPEGTKFTF- 982
y 1150 SCEPSSHDTNSV-----LNTSGNSISGTVPLKRRPEDDE 1185
b 983 -TEINYEKDNVNGKHNEDEKESQTLTPKEVETIPSTPKQPE 1024

```

```

; RESULT 7
; JS-09-252-991A-16913
; Sequence 16913, Application US/09252991A
; Patent No. 6551795
; ORGANISM: Streptomyces aeruginosa
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190

```

```

; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16913
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-16913

Query Match
Best Local Similarity 1.9%; Score 118.5; DB 4; Length 370;
Best Local Similarity 20.2%; Pred. No. 0.0051;
Matches 93; Conservative 46; Mismatches 137; Indels 185; Gaps 19;

QY 208 POPFNTHEISGEGFYPTFFNTGLCHLDVCPMAVALRVENVTAVARGAAHLAFD 267
Db 17 PDCNAYH-----TDEQWAAVLTFD 37
QY 268 ENHEGA-VLPDDITYTFQSSSGTTTARGARNVDN--STSKPSPSGGF-----ER 316
Db 38 AAADGAFVAVKTTGVYCRPSSS---ARRPRNVEFPFATAEAAAGYRPSRRAAGDR 93
QY 317 RLASINAAATLHAEIFNTGIYEBETPTDIKEMPMFIGME-----GTLPRLNA 364
Db 94 RLAAEQRAERVQAACHMIETA---ETPPALEALAAARLGMSPPHFRLFKAETGLTPKAYA 150
QY 365 LGSYTRAVAGVIGAMVFPSPNSALYLTVEDEDSGWTAKDGGPGSPFNRYQFAGPHLAANP 424
Db 151 SAYRARELRERLQASASVTEAIY-----DSGFN-----SNSRFEVSSQRLGMRP 196
QY 425 QTRDRGHVLSQSSTGSGSNTSFVDYALICGAGPALLARLLFYLERCDAGFTGGHGDA 484
Db 197 RDYRDG-----GAG-----AAIRFAIQCCSLGAI----- 220
QY 485 KYVTGTFDSEIPCSLCEKHTRPVCA-----HTTVHRLRQSMR-----EQOATR 528
Db 221 -----LVAQSRGICAILLGEPEPILLRELOQFPRAQLLGGDADFERLVA 266
QY 529 QPIGVFTGWNQSDCPGNYAPYILIKPKGQTEAAKATMQ-----DTVRATLERLF 582
Db 267 QVVG-----FVESPOLGLDLP-LDVRTGTAQFQVQALREIIPPGSTASYAQIAERI- 316
QY 583 IDLEQERLLDRGAP-----CSSEGLSSVVDHPTFRR 614
Db 317 -----GAPRAVRAVAQAACANRIAPVPCRVVRR 346

```

```

; RESULT 8
; US-09-105-537-2
; Sequence 2, Application US/09105537A
; Patent No. 6265202
; ORGANISM: Streptomyces venezuelae
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 5215
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
; US-09-105-537-2

Query Match
Best Local Similarity 1.8%; Score 111; DB 3; Length 5215;
Best Local Similarity 19.7%; Pred. No. 4.4;
Matches 250; Conservative 109; Mismatches 419; Indels 494; Gaps 66;

```

```

QY 51 NLTKETFTTSSLAV-----SGARTGLAGAGTTLKLTSHFYPSVFPVHGKGVLP 102
Db 166 SLTVDAACQSSSLVAVHLACBSLRAGESTTALV-AGVNL-----NILA 206

```

103	QY	SSAAPNLTRACNAARRRFRGSPRCQGPVGDAGVETTGAEICTRLGLBPEPNTILVVLVUTALF	162
207	DB	ESA-----VTEBFFG-----GLSPDGTAVTFDARA-----	231
163	QY	KEAVFNCVFLHYGGDLIVHINHGVDVIRIPLFPVQLFMPDYNRLVPDPENTHRSIGEGF	222
232	DB	-----NGEVREG-----GGVVVKEL-----SALADGDRVH-----GV	261
223	QY	VYPTPTNTGLCHLIHDCVIAPNAVALRVNRVTVAVARGAAHLAFDNHGGAVLPDPITVT	282
262	DB	IRASAVNNDG-----ATPGLVPSRAAQEKVLRBAVYKAAALDPS-AVQ	303
283	QY	YFOSSSGTITARGARRNDVNSTKSPSCGFERRLASIMAAADTALHAZVIFWTGYEET	342
304	DB	YVELHGTGT-----PVGDPTEAALGAVIGSAPADBPJ-----LVGSA	342
343	QY	PTDICKEMPFMEIGETLPRINALGSYTVARVAGVIGAMVFSNSALYLTEYEDSGMTEAKD	402
343	DB	KTNVGHLEGAAGIVGLIKTLALG-RRRIPASLNPRTPHPDIPJ-----DTLGLDVP	394
403	QY	GGPGPSFNRYQFAGPHLANPQTDPR-----GHVLSQSQTGSSNTFFSVD	448
395	DB	G-----LREWPHDRELLAGVSFGMGSTNAHVLSBGPAGQGEQFGID	438
449	QY	YLALICGFAPLLARLLFYLERCDAGA-----FTGGHGDAK-----YVTGTFDSEI-P	496
439	DB	BEPTV-----DSNALFPVVTGRGEALRAQAARLHEAVEADPELAP	480
497	QY	CSICEK-HTRPYCAHTV-----HRLORMPRFGQATQPTQIVGFTMMSYQSDCDPLG	548
481	DB	AALARSIVTTRTVFTHRSVVLAPDRARLLDGLGALAAGTAPGVTGT-----	528
549	QY	NYAPYLLRLRPGQTEAAKATMQDTYATLERLFDLEQERLLDRGAPCSGESLSVIVD	608
529	DB	-----PAPG-----RLAVLF-----SGQGRQRTGMGV-ELYAA	555
609	QY	HPTRFRILTRARIEQTTTFQMKVLVETRD-----YKIREGJSEATHSNALT	656
556	DB	HPAFATAFDAVAELEDPLDRPLAEALVAAGDTLDRTVHTQPALFAVEVALHRLVSWGVT	615
657	QY	PDYSGAFCSITNFKRTHLAVQDIALSQCHCVYGOQVEGFENFRNQPVLRFRFVD	716
616	DB	PDLIAGH-----SVGEISAAHVAGY-LSLRDA-----ARLVAARGRLMQALP	656
717	QY	LFNGGFSTRSITVTLSEGFVSAPNPTLQGDAPAGRTFGDGLARVSVEVIRDVRKRVV	776
657	DB	-----EGGAM-----VAVEASEEVL-----PHL-----AGREREISLAVN-----GPRAVV	694
777	QY	PSGCTNLSEAAARLRYGLASAYORQEKRVDMHLHGALGFLIKPHGILFF	825
695	DB	LAG-----ABRAVLDAELLREQGRRTKRLSVSHA-----FHSPLMBPMLDDFRVV	741
826	QY	-----PR-----GMPNSKSPFOW-----FWTLLQRNQNPADKLTHEEITIAAV	866
742	DB	EELDFQBPVRDVSIVTGLPVT-----GQMTDPEY-----VDQV-RRPVRFIDAV	787
867	QY	KRFTEEVAAINFINLPP--TCIGELAQFYMANILIKYCDHSQYLINTLTISITGAERPD	924
788	DB	-RTLEESGADTFLELPGDGVCSMAADSV-----RDQEAATAVSALRKG--RP-E	833
925	QY	PSSVLHWTRKQDVTSAADIETOAKALLEKTEMLPELWTTAFTSTHLVRAAMNQRPVVLGI	984
834	DB	QSLAALATTVFVRGHVD-----WTAAGHSTGTVRVPL-----PTYAFOR	874
985	QY	SISKYHGAAGNRRVFOAGNWSGLGGKNCVPLFTFDRTRFRPIACPRGGFCPTVGPSSG	1044
875	DB	ERHWFDDGAARTAAPLTAGR-SGTGAG-----	905
1045	QY	-----NRETTLSQDVRGIIVSGGAMVOLAIYATVVRA-VGARA	1081
906	DB	VTSGEGEGBGAGGAGGDRPARHETT--ERVRAHVA-----VLEYDDPTRVLELGTFF	957

```

1283 -----ARGRSCTSVAMTPWALPGAVDDGYLRRGLRS-LSADRAMTWERVLA 1330
336 TGIYETPTDIKEWPMFI-GMEGTLPRLNALGSYARVAGVFMVFNPSALYITEVED 394
1331 AGPVSAVADV-DWPLVSEGFATR------TALFAELAGR 1365
395 SGWTEAKDGGPSPFNRFYQFAG------PHLAANPQTDROGHVLSOSTGSSNTE 444
1366 GGOAEAPDS-GTGPBPAQLAGLSPDEOENLELVANAVAE------VLGHSAABINVR 1420
445 FSDYDALICGFGAPLRLLFYLERCDAGATGGHGDALKYVTGTDFSEIPCSICEKHT 504
1421 RAPSGLDLSINAMALRKL------SASTG-----LRLPASL----- 1452
505 RPYCAHTTWHRLRQ--RMPRFQATROPICVFGTMSQVSDCDPL-----GNVA 551
1453 --VFDHPTVTALAOHLRLARLVGDQAARVAVGAAD-----ESEPIAIVGICRFPFGIGS 1506
552 PYLILR-----KPGDOTEAA----- 566
1507 PEQLWRVLAEGANLTGFPADRGWDIGRLYHPDPNPGTSYVDKGGFLTDADFPDGFPG 1566
567 -----KATMDTYRATLERFLDLQERLLDRGAPCSSEGSSVIVDHTFRR 614
1567 ITPREALAMDPOORLMLTAWEAVERAGIDPDALRGTDGTGVFMNGOSYMOLLAGEAER 1626
615 I-----LDTLRARIET-----TTQFMKVLAVETRDYKIREGLSEATHSWAL 655
1627 VDGYOGLNSASVLSGRIATYTFWEPAUTVDTCSSLVGI--HLAMQALRGCSLAL 1684
656 T-----FDPYSAGFCPTNFVLRKTHLAVVDDALSOCHVFGYQQVQVEGRNFRN----- 704
1685 AGCVTVMSDPYT-----FVDFSTQR-----CLA-SDGRCKAFSARADGFLSSEVAAL 1731
705 QPQVLRPRFVDFLNGGFISTRSITVTLSEGPVSAPNPTLQODAPAG-----RTFPGDLAR 760
1732 VLEPLSRAR-----ANG-----HQVLAVRGSVAVNQGASGLAAPNGSPQERVIQALAA 1782
761 VSEVIRDIRVKNRVFVSGNCTNLSEAAARLVGLASAYORQEK------VMDLH-- 810
1783 SGVPA-ADVDV---VBAHGTGTGLGPIEAG--ALITVGGDRDRPLRGLSVKTHIGHTQ 1836
811 -----GALGFLKQPHGLIFRGMPPNSKSNPQWFTLLQNNQMPADKLTHEBITTIA 864
1837 AAAGAAGVIVKVLAMRHGL-PSLHADELSPHDW-----ESGAVEVLRSEVPWPA 1887
865 AVKRFTEVAAINFIMLPPTCIGELAQFWMANILKYCDHSQVYLINTLTSITIGARRPD 924
1888 GER---PRRAGVSSFGVSGT-----NAHVIVEEAPAEQ-----EAARTERG 1925
925 PSSVLHWIRKDVTSAADITQAKALLEKTENLPEL-----WTTAFTSTHL-VRAAMNQR 977
1926 PLFPVLSGRSEAVAA---QARALAEHLRDPPELGLTDAAMTLATGRARFVRAV--- 1978
978 PMVVLGISIKYHGAAGNRRVFOAGNWSGLNGKNCVPLFTFDRTRFRFTIACPRGFCIP 1037
1979 -----LQDRAGVCAELDALAER-----PSADAVAP 2005
1038 VTGPSSGNRETTILSDQVRGIIVSGGAMVQLAIYATVAVRAGARQAQHWAFDDLSLTDDEF 1097
2006 VT---SAPRXPVL-----VFPQGAQ-----WVG----- 2026
1098 LARDLLEHLDQIQTLL-----ETPWTVEGALBAVKILDEKTTAGDG 1138
2027 MARDLLESSEVFAESMRCAEALSPTHDM-----KLLD--VVRGOG 2065

```

RESULT 10
 US-08-439-009A-4
 ; Sequence 4, Application US/08439009A
 ; Patent No. 6004787
 ; GENERAL INFORMATION:

```

; APPLICANT: Donadio, S
; APPLICANT: Katz, L
; APPLICANT: McAlpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; TITLE OF INVENTION: Specific Polyketides
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven F. Weinstein
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; STREET: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/439,009A
; FILING DATE: 11-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Casuto, Dianne
; REGISTRATION NUMBER: 40,943
; REFERENCE/DOCKET NUMBER: 4952.US.D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847-938-3137
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3567 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-439-009A-4

```

```

Query Match 1.7%; Score 106.5; DB 3; Length 3567;
Best Local Similarity 19.9%; Pred. No. 6.2;
Matches 237; Conservative 104; Mismatches 384; Indels 465; Gaps 62;

QY 125 CQGPVVDGAVETTGAE--ICTR-----LGLPEPTILYLVVVTALFKEAVFMCNVPLH 174
DB 1165 CPGDLLAAVEEAGASAVVCAQAAALRALGDEP-----VTAL-----VH 1205
QY 175 YGGLDIVHNGDVIRIPLFPVQLFMPDVNRILVPDPFNTHRSIGEGFYVPTFPYNTGLC 234
DB 1206 AGTL-----TNFG-----SISEVAPEEF-----AETIAAKTALL 1234
QY 235 HLHDCVIAPMAVALRV-----RNVAVARGAAHL-AFDENHEGAVLPDITVT 282
DB 1235 AVL-DEVLGDRVAREYCYSSVAGIWWGAGMAAAYAGSAYLDALAEHR----- 1282
QY 283 YFQSSSGTITAGARRNDVNSTKPSQG---GF--ERRLASIMAAADTALHA-EVIFN 335
DB 1283 -----AGRSCTSVAMTPWALPGAVDDGYLRRGLRS-LSADRAMTWERVLA 1330
QY 336 TGIYETPTDIKEWPMFI-GMEGTLPRLNALGSYARVAGVFMVFNPSALYITEVED 394
DB 1331 AGPVSAVADV-DWPLVSEGFATR------TALFAELAGR 1365
QY 395 SGWTEAKDGGPSPFNRFYQFAG------PHLAANPQTDROGHVLSOSTGSSNTE 444
DB 1366 GGOAEAPDS-GTGPBPAQLAGLSPDEOENLELVANAVAE------VLGHSAABINVR 1420
QY 445 FSDYDALICGFGAPLRLLFYLERCDAGATGGHGDALKYVTGTDFSEIPCSICEKHT 504
DB 1421 RAPSGLDLSINAMALRKL------SASTG-----LRLPASL----- 1452
QY 505 RPYCAHTTWHRLRQ--RMPRFQATROPICVFGTMSQVSDCDPL-----GNVA 551
DB 1453 --VFDHPTVTALAOHLRLARLVGDQAARVAVGAAD-----ESEPIAIVGICRFPFGIGS 1506

```

552 PYLIR-----KPGDQTEAA----- 566
1507 PEQLWRVLAEGANLTTGPPADRGWDIGRLYHPDPNPGTSYVDKGGFLTDAAFDPGFFG 1566
567 -----KATMODTYRATLERLIDLEQERLLDRGAPCSSESGLSSVIVDHPTEFR 614
1567 ITPREALAMDQOORLMLETAWEAVERAGIDPDALRGDTGVFGVMNGQSYWQLLAGEAER 1626
615 I-----LDTLRARIET-----TTFPMKVLVETRDYKIREGLSEATHSMAL 655
1627 VDGQGLGNSAVLSGRITAYFGWEGPALTVDTACSSSLVGI--HLAMQALRRGECSLAL 1684
656 T-----PDPYSGAFCPITNLFVKRTHLAVVQDLALSOCHQVFGQOQVEGRNFRN----- 704
1695 AGVTVMSDPT-----FVDFSTGR-----GLA--SDGRCKAFSARADGAFALSEGVAAL 1731
705 QFQVLRFRFVLDLNGFGFISTRSITVTLSEGVAPNPTLGGDAPAG-----RTFDGLAR 760
1732 VLEPLSRAR-----ANG-----HOVLAVLRGSVAVNQDASGLAAPGPSQBRVIRQALAA 1782
761 VSVEVIRDIRVKNRVVFGSGNCTNLSEAAARLVGLASAYQOEKR-----VDMHLH-- 810
1783 SGVPA-ADVAV--VEAHGTGTGLGDPLEAG--ALIAITYGQDRDRPLRLGSKVTNIGHTQ 1836
811 -----GALGFLKQFHGLLFPGRMPNPKSPNQWFMTLLQRNQMPADKLTHEEITIA 864
1837 AAAGAAGVKKVLMRHMGL-PRSLHADELSPHIDM-----ESGAVEVLRVEVPWA 1887
865 AVKRETEYAINFNLPTTCIGELAQFYMANLKYCDHSOYLNTLTSITTOGARPRD 924
1888 GER-----PRAGVSSFGVSGT-----NAHVIVEEAPAEQ-----EARTERG 1925
925 PSSVLHWIRKQVTSADTETQAKALLEKTENLPEL-----WTTAF*STHL-VEAAMNOR 977
1926 PUPFVLSGRSAVAA-----QARALAEHLRTPELGLTDAWNTLATGRAPDVRAAV-- 1978
978 PMVLGISISKYHGAAGNNRVFQAGNWSGLNGKKNVCLFTFDRTRRFFIACPRGGFICP 1037
1979 -----LGDRAGVCAELDALAEGR-----PSADAVAP 2005
1038 VTGPPSGNRETTLSDOVRGIIVSGGAMVQALAIYATVRAVGARAHQWAFDDMLSLITDEF 1097
2006 VT---SAPRKPVL-----VFFGQGAQ-----WVG----- 2026
1098 LARDEELHQDIQTL-----ETPWTVEGALEAVKILDEKTTAGDG 1138
2027 NARDLLESEVEFAESMSRCAEALSHTDM-----KLID--VVRGDG 2065

RESULT 11
US-07-731-157A-7
Sequence 7, Application US/07731157A
Patent No. 5457032
GENERAL INFORMATION:
APPLICANT: Quax, Wilhelmus J.
APPLICANT: Misset, Onno
APPLICANT: Van der Laan, Jan M.
APPLICANT: Lenting, Hexman B.M.
TITLE OF INVENTION: Mutated beta-lactam acylase genes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESSES:
ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/731.157A
FILING DATE: 19910509
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 90200962
FILING DATE: 18-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: RAE-VENTER PH.D., BARBARA
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: GERO-027/00US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-494-7622
TELEFAX: 415-857-0663
TELEX: 380816 COOLEY PA
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 774 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Pseudomonas species
STRAIN: SE83
US-07-731-157A-7
Query Match 1.7%; Score 106; DB 1; Length 774;
Best Local Similarity 20.2%; Pred. No. 0.4;
Matches 145; Conservative 79; Mismatches 251; Indels 242; Gaps 37;
QY 258 ARGAHLAFDENHEGAV-LPDDITYTFQSSSGTTRTARGARRNDVNSTKPSPSGGPER 316
DB 183 AANALKRYDDGGQDLGICPGVEAELEADLAALRPAVDALLKAMGDADAAGGSNN 242
QY 317 -----RLAS-----IMAAATLHAEVIENTGVEETPTDIKEWPMFIMGEGTLPRNLALGS 367
DB 243 NAVAQRTRATGRPILAGDPHRVFEI-----PGMYAQHFLACDRFDM-IGL--TVP----- 289
QY 368 YTARVAGVIGANVFPSNA-----LYLTVESDGMTE----- 399
DB 290 -----GVDPGFPHAHNGKAVCVTHAFWDIHDLYLEQFAEDGRTARFQNEPEPVAMRRD 343
QY 400 --AKDGGPSPNRFPYQAGPHLAANPQTDGRGHVLSQSQTSSNTSFSDVYALICGFG 457
DB 344 RIARVGGADREFDIVETRHGPFVIAGDP--LEGAATLRSVQFAETDLSFDCLTNP--G 398
QY 458 APILARLLFYLERCDAGAFTG--GH---GDALKYVTGTFTDSEIIPCSLCEKHTRVVCART 511
DB 399 ASTVAQLY-----DATRGWGLIDHNLVAGDVAGSIGHLVRAVFSRPRENGMLPVPGWS 452
QY 512 TVHRLR-----QRMPRFGQATEQPIGVFGTMNSQY-----SDCDPLGNVAPY---- 553
DB 453 GEHEWGRGMPHEAMPR--VIDPPGGLIVTANNRVVADDPDYLCTDCHP-----PYRAE 504
QY 554 -----LILRKGQDTEAAKATMODT-----VRATLERLFI--DLEQERL----- 590
DB 505 RIMERLVASFAFAVDDAAIAHADTSLSPHVLRLARLEALIGQSLPAEELRQTLIAWDGR 564
QY 591 LDRGAPCSSEGLSSVIVDHPTRRILDTLRLARIEQTTFQMKVLVETRDYKIREGLSEA- 649
DB 565 MDAGSQAASA-----YNAFRRAL-----TRLVTARSQLEQAI 596
QY 650 THSNALT-----FDPYSGAFCPITNLFVKRTHLAVVQ-----DLALSOCHQVFGQOQVEGRNF 702
DB 597 AHPFAAVFPFGVSPGQGVWAVPT--LLRNDADAGLKGWSWDEALSEALS--ATQNLITGRGW 654
QY 703 RNQFPQVLRFRFVLDLNGFGFISTRSITVTLSEGVAP-----NPTLQODAP 749
DB 655 GEEHRP-----RTHPLSAQFPAAWALL-----NPVSRPIGGDGTVLANGLVPSAGPEAT 705
QY 750 AGRTFQDGLARVSEVIRDIRVKNRVVFGS--GNCNTNLSEAAARLVGLASAYQOEKRVDM 808
DB 706 YG-----ALSRYVFDVGNWDN-----SRWVV 726

1Y 809 LHGALGFLKQFHGLLFRPGMPPNSKSNPQW-----FWTLQRNOMPADKLTHEEI 860
1b 727 FHGASG-----HPASPHYADQNAPMSDCAMVPMLSYMDRIAIAEAVTSQEL 771

RESULT 12
S-08-541-780-7
Sequence 7, Application US/08541780
Patent No. 5935831
GENERAL INFORMATION:
APPLICANT: Quax, Wilhelmus J.
APPLICANT: Misset, Onno
APPLICANT: Van der Laan, Jan M.
APPLICANT: Lenting, Herman B.M.
TITLE OF INVENTION: Mutated beta-lactam acylase genes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
STREET: COOLEY GODWARD CASTRO HUDDLESON & TATUM
CITY: FIVE PALO ALTO SQUARE, 4TH FLOOR
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94306

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/541,780
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/731,157
FILING DATE:

APPLICATION NUMBER: EP 90200962
FILING DATE: 18-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: RAE-VENTER PH.D., BARBARA
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: GERO-027/00US
TELEPHONE: 415-494-7622
TELEFAX: 415-857-0663
TELEX: 380816 COOLEY PA

INFORMATION FOR SEQ ID NO. 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 774 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Pseudomonas species
STRAIN: SE83

IS-08-541-780-7

Query Match 1.7%; Score 106; DB 2; Length 774;
Best Local Similarity 20.2%; Pred. No. 0.4;
Matches 145; Conservative 79; Mismatches 251; Indels 242; Gaps 37;
2Y 258 ARGAAHLAFDENHGAVALPPDITYTFSSSGTTTARGARRNDVNSTKPSGSGFER 316
1b 183 AANALXRYDDGQDLCIPPGVEARLEADLAALRPVADALLKAMGGDASDAAGGSGNN 242
2Y 317 -----RLAS--IVAADTALHAEVFNFTGIYEETPTDIKEMPFEGMTEGLPRINALGS 367
1b 243 WAPVORTATGRPLAGDHRVFEI---PQWQAHLACDRDM-IGL--IVP----- 289
2Y 368 YVARVAGVIGAMVFPNSA-----LYLVEVDSGMT----- 399
1b 290 -----GVGFFHFAHNGKVCVTHAFWDIHDLYLEQFAEDGRTARFGNEPEPVAMRDR 343

QY 400 --AKDGGPSPENRFFQFAGPHLAANPQDTRDCHVLSQSTGSSNTEFSVDYLALICGFG 457
Db 344 RIARVARGAREFDIVETRHPVIAQDP---LEGAALTLSRVQFAETDLSFDCLTXP--G 398
QY 458 APLLARLLFVLERCDAGAFG--GH-----GDALKVVTGTFTDSEIIPCSCLEKHTRVCAHT 511
Db 399 ASTVAQLY-----DATRCWGLIDHNLVAGDVAGSIGHLVRAVPSRPRENGWLVPCWS 452
QY 512 TVHRLR-----QRMPPFGOATQPIGVFGTWNQY-----SDCDPLGNVAPY--- 553
Db 453 GEHEMRGWIPEAMPR---VIDPPGGLIVTANNRVVADDDHPDYLCCTDCHP-----PYRAE 504
QY 554 ----LILRKPGQOTAAKATMODT-----YRATLERLFI--DLEQERL----- 590
Db 505 RIMERLVASPAFVADDAAHADTSLSPHVGLLRARLEALGIQGLPAELRQTLIANQGR 564
QY 591 LDRGAPCSSEGLSSVIIVDHTFRRIDTLRARIEQTTTFQMKVLVETRDYKIREGLSEA- 649
Db 565 MDAGSQAASA-----YNAFRAL-----TRLVTARSGLQAI 596
QY 650 THSMALT---FDPYSGAFCPITNPLVKRTHLAVQ---DLALSQCHCVFYGQVQYGRNF 702
Db 597 AHPFAAVPFGVSPQGVVWAVPT--LLRNDGAGMLKGNWDEALSEALSV-ATQNTLGRGW 654
QY 703 RNQFQVLRERRFVDFLNFNGGFISTRSITVTLSGPPVSAP-----NFTLGQDAP 749
Db 655 GEEHRP-----RTHPLSAQFPAAWALL-----NPVSRPIGGDGTVLANGLVPSAGPEAT 705
QY 750 AGRTFGDGLARVSEVIRIRVKNRVVPS-GNCTNLSEARARLVGLASAYQQRKRVDM 808
Db 706 YG-----ALSRYVFDVGNWMDN-----SRWVY 726
QY 809 LHGALGFLKQFHGLLFRPGMPPNSKSNPQW-----FWTLQRNOMPADKLTHEEI 860
Db 727 FHGASG-----HPASPHYADQNAPMSDCAMVPMLSYMDRIAIAEAVTSQEL 771

RESULT 13

US-09-107-532A-3855
Sequence 3855, Application US/09107532A
Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277
 INFORMATION FOR SEQ ID NO: 3855:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1095 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: YES
 ORIGINAL SOURCE:
 ORGANISM: Enterococcus faecium
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (B) LOCATION 1..1095
 SEQUENCE DESCRIPTION: SEQ ID NO: 3855:
 US-09-107-532A-3855

Query Match 1.7%; Score 105.5; DB 4; Length 1095;
 Best Local Similarity 18.9%; Pred. No. 0.86;
 Matches 145; Conservative 112; Mismatches 298; Indels 211; Gaps 35;

551 APVILLRKPGDTEAAKATWQDYRATLERFLDLEQRLLDGAPCSSEGLS-----S 604
 360 APTIDTPTWTTIKAGETIVTSKNAQKQILDKSG-VETGSLMNDNYSLAGNTFA 418
 605 VIVDHPFRRLD-TLRARIEQTTQPMKVLVETRDYKIREGLSEATHSWALTFDP--- 659
 419 IRKDSPTGEIVQEMTTDENGHAETPREIANALELGIYYVE--TKASHGFVNTFKVKVE 476
 660 --YSGAFCPITNLFVKRTHLVVQDLAL-----SOCHCVFYGOQV-----EGR- 700
 477 LKVANOTVALVTSNVKQNCQEVGTETTLTKEDKDTGDKAGKAVFEGTGYTLFTAKDGKA 536
 701 -NPRNOPQVLRFRFVDFLENGGISTRSITVTLSEGVSAF-----NPTLGD--APAGR 752
 537 VKSEAFKP-----EMVGTQASDETVTLADENKQAAVKHLAINFYWGETRAPEGY 589
 753 TFGDGLARVSVE-----VIRDVRKRVV-----FSGNCTNLSBA----- 788
 590 TLDETKVPVSIKKVDNEKNNAVITRDVTAKEIIRFGDFPKFAGSAAGTAETGFNDLTF 649
 789 -RARVLGLASAYQKQKRVMDLHGAIGF-----LLKQFHGLLPRGM--- 829
 650 KVPFLEGTNEITGAEDTATYNEQGLGDFGKFNLPYGDYLLLEEVA---PEGFQKIT 706
 830 PPNKSKSP-----NPQFWFTLLQRNQ-----MPADKLTHEEITIAAVKRFTEE 872
 707 PLEIRSTFKENKEDPVKSEYVFVITEQDQKPIKTVTPVEKLTN-----KAFSVS 757
 873 YAINFINLP--PTCIGELAQFYMANILKYCDHSQVYLINTLSITTGARRPRDPSSVLH 930
 758 LNRLMLYDLPEEEDSLTSLATWKDGNKELTSLDSTE-LVDKL-----SYNLH 803
 931 WIRKD--VTSADIEQAKALLEKTENLPELWTTAFTSTHLVRAAMN-----QRPVW-- 980
 804 EIKEDVWVQAQIDVDA-TKAAQEKDEKAPV-VIAETSLANKETGKWKIQLKLTAE 861
 981 -VIGISISKYHGAAGNNRVFQAGNWSGLNGKNCVCLFPTDRTRFIIACPRGGFICPVY 1039
 862 QVLNKTIVLENYVENKEAFEAG-----KQVA 889
 1040 GPSSGNRETTLSQVRIIYSGGMVQALAIYATVAVRAGARQAQMAFDWLSLTDDEFLA 1099
 890 -----KDVSLNNAQVTSCTVEHHVSIQTKAHLENG-----SQTFTHGVDVDFDVSIT 939
 1100 RDL-----EELHQDIQIOTLETPWTVEGALEAVKILDE-----KTT-----AGGETPTNL 1144
 940 HDVLDSGKEAFETILYALLPDGNTKEWKSGKIFYEVNDKEFTKTVLAKVDYDKYKEPT 999
 1145 AFNTDSCPSHDTTSN-----VLNISGSNIGSTVPGCLKRPDPDE 1185
 1000 KFTFAEINYDKDGTINGKHNEDLKEKSQTLTPKEVPTILTSPKQPE 1045

RESULT 14
 US-08-633-760-46
 ; Sequence 46, Application US/08633760
 ; Patent No. 5804429
 ; GENERAL INFORMATION:
 ; APPLICANT: NIWA, MINEO
 ; APPLICANT: SAITO, YOSHIMASA
 ; APPLICANT: FUJIMURA, TAKAO
 ; APPLICANT: ISHII, YOSHINORI
 ; APPLICANT: NOGUCHI, YUJI
 ; TITLE OF INVENTION: A NEW CEPHALOSPORIN C ACYLASE
 ; NUMBER OF SEQUENCES: 64
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
 ; ADDRESSEE: P.C.
 ; STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE 400
 ; CITY: ARLINGTON
 ; STATE: VIRGINIA
 ; COUNTRY: USA
 ; ZIP: 22202
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/633,760
 ; FILING DATE: 01-MAY-1996
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: OBLON, NORMAN F.
 ; REGISTRATION NUMBER: 24,618
 ; REFERENCE/DOCKET NUMBER: 18-929-0 PCT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 413-3000
 ; TELEFAX: (703) 413-2220
 ; TELEX: 248855 OPAT UR
 ; INFORMATION FOR SEQ ID NO: 46:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 774 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-633-760-46

Query Match 1.7%; Score 105; DB 1; Length 774;
 Best Local Similarity 20.7%; Pred. No. 0.5;
 Matches 146; Conservative 76; Mismatches 241; Indels 242; Gaps 37;

QY 109 LTR--ACNAARERFGSRCCQPPVDGAVETTGAE-ICTR-----LGLPEPTILYLVVTAL 161
 DB 70 LTRRKALGRAAEMLG---ABAAEADILVRLGMEKVCRRDFEALGVEAKD-----M 117
 QY 162 FKEAVFNCVFLHVGGLDIVHINHGVIR:PLPVPVOLFMPDVRNLVDPDPTNTH-----R 216
 DB 118 LRAVAGVNAFLASGA-----PL-PVEYGLIGAE---PEPWPWHSIAVMR 159
 QY 217 SIGGFVYPTFFNTYGLCHLHDCVIAAPVALRVNRVTAV----ARGAAHLAPDENHEG 272
 DB 160 RLG-----LLMGAVWFKLWRMLALPVVGAANALKRLRYDDGGRD 197
 QY 273 AV-LPPDITTYTFOSSSGTITARGARNVDNSTSKSPSGGFER-----RLAS----IM 322
 DB 198 LCLTPGAEDRLAEADLALRPAVDALLKAMGGDASDAAGGNNNMAVAGRTATGPIL 257
 QY 323 AADTALHAEIFNTGIYEETPTDIKEWPMFIEGEGTLPRLNALGYSYARVAGVIGAMVFS 382
 DB 258 AGDHRVFEI---PGMYAQHHLACDREDM-IGL--TYP-----GVPGPQHPFA 298
 QY 383 PNSA-----LYLTEVEDSGMTE-----AKDGGPPSPSNRF 412
 DB 299 HNGKVAICVTHAFMDIHDLYLEQFAGEGRTARFGNDPEPVAWSRDIARVGGADREDFIV 358

QY 413 YQAPGHLAANPOTDRDGHVLSQSTGSSNTSPESVDYLALICGFGAPLLARLLFVLERCD 472
 Db 359 ETHGPIVAGDP---RGAALTLSRVQFAETDLSFCLTEMP--GASTVAQLY-----D 407
 QY 473 AGAFTG--GH-----GDALKVVTGTFDSEIPCSICEKHTRPVCAHTTVHRLR-----QMP 521
 Db 408 ATRGWLIDHNLVAGDVAGSIGHLVRAVPSRPRENGWLPVPGWSGEHWRGWIPIHEAMP 467
 QY 522 RFGQATRPQIGVGTWNSQY-----SDCDPLGNVAPY-----LILKPGDQTE 564
 Db 468 R---VIDPPGGIIVTANNRVVADHDHFDLCTDCHP-----PYRAERIMKRLVANPAFAVD 519
 QY 565 AAKATMDQTVRATLERLFIIDLEQERLDRCA--PCSEGLSSVIV-----D 608
 Db 520 DAAAIHADTLSP-----HVGLRRRLREALGARDSDAAGLRQMLVANDGRMDAASEVASA 574
 QY 609 HPTFRILDTL---RARIETTTQ-----FMKVLVETRDYKIREG--- 645
 Db 575 YNAFRALTRLVTRDSGLEQAISHPFAAVPGVSPQGVWMAVPTLLRDDDAGMLKGWSW 634
 QY 646 ---LSEA-----THSMALTDPYSGAFCP 666
 Db 635 DQALSEALSVASQNLTCRSWGEHRRPFTHTPLATQPPAWAGLLNP 679

RESULT 15
 JS-08-633-760-48
 ; Sequence 48, Application US/08633760
 ; Patent No. 5804429
 ; GENERAL INFORMATION:
 ; APPLICANT: NIWA, MINBO
 ; APPLICANT: SAITO, YOSHIMASA
 ; APPLICANT: FUJIMURA, TAKAO
 ; APPLICANT: ISHII, YOSHINORI
 ; APPLICANT: NOGUCHI, YUJI
 ; TITLE OF INVENTION: A NEW CEPHALOSPORIN C ACYLASE
 ; NUMBER OF SEQUENCES: 64
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
 ; STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE 400
 ; CITY: ARLINGTON
 ; STATE: VIRGINIA
 ; COUNTRY: USA
 ; ZIP: 22202
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/633,760
 ; FILING DATE: 01-MAY-1996
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: OBLON, NORMAN F.
 ; REGISTRATION NUMBER: 24,618
 ; REFERENCE/DOCKET NUMBER: 18-929-0 PCT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 413-3000
 ; TELEFAX: (703) 413-2220
 ; TELEX: 248855 OPAT UR
 ; INFORMATION FOR SEQ ID NO: 48:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 774 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; JS-08-633-760-48

Query Match 1.7%; Score 105; DB 1; Length 774;
 Best Local Similarity 20.7%; Pred. No. 0.5;
 Matches 146; Conservative 76; Mismatches 241; Indels 242; Gaps 37;

QY 109 LTR--ACNAARERFGRSCOGPPVDGAVETTGAB-ICTR-----LGLEPENTILYLVVTAL 161
 Db 70 LTRKALGRAAEMLG---ABAAEADILVRLGMEKVCRRDFEALGVEAKD-----M 117
 QY 162 PKEAVFMCNVLHVGGLDIVHINHGDVIRIPLFPVOLFMPDVNRLVDPDPNTHH-----R 216
 Db 118 LRAIVAGVNAFLASGA-----PL-PVYGLLGAE---PEWEPWHSIAVNR 159
 QY 217 SIGEGFYPTFPYNTGLCHLIDCVIAPMAVALRVNRVTAV-----ARGAAHLAFDENHEG 272
 Db 160 RLQ-----LIMGSVWFKLWRMLALPVVGAANALKRYDDGGRD 197
 QY 273 AV-LPPDITTYTCSSSSGTTTARGARRNDVNSTKPSGSGFER-----RLAS---IM 322
 Db 198 LLCPPGAEDRLEADLTLRPAVDALLKAWGDSDAAGGSSNNWAVAPGRATGRPL 257
 QY 323 AADTALHAEIFNTGIYEETPTDIKEWPMFIMEGLPLRLNALGYSYTVARVIGAMVFS 382
 Db 258 AGDPRHYFEI---PGIYAQHHLACDRFDM-IGL--IVP-----GVPGFPHFA 298
 QY 383 PMSA-----LYLTVESDGMTE-----AKDGGPGSPNRF 412
 Db 299 HNGKVAYCVTHAFMDIHDLYLEQFAGEGRTAREGNDFFPVAWSRDRIVRGGADREFDIV 358
 QY 413 YQAPGHLAANPOTDRDGHVLSQSTGSSNTSPESVDYLALICGFGAPLLARLLFVLERCD 472
 Db 359 KTRHGPVIAGDP---RDGAALTLSRVQFAETDLSFCLTEMP--GASTVAQLY-----D 407
 QY 473 AGAFTG--GH-----GDALKVVTGTFDSEIPCSICEKHTRPVCAHTTVHRLR-----QMP 521
 Db 408 ATRGWLIDHNLVAGDVAGSIGHLVRAVPSRPRENGWLPVPGWSGEHWRGWIPIHEAMP 467
 QY 522 RFGQATRPQIGVGTWNSQY-----SDCDPLGNVAPY-----LILKPGDQTE 564
 Db 468 R---VIDPPGGIIVTANNRVVADHDHFDLCTDCHP-----PYRAERIMKRLVANPAFAVD 519
 QY 565 AAKATMDQTVRATLERLFIIDLEQERLDRCA--PCSEGLSSVIV-----D 608
 Db 520 DAAAIHADTLSP-----HVGLRRRLREALGARDSDAAGLRQMLVANDGRMDAASEVASA 574
 QY 609 HPTFRILDTL---RARIETTTQ-----FMKVLVETRDYKIREG--- 645
 Db 575 YNAFRALTRLVTRDSGLEQAISHPFAAVPGVSPQGVWMAVPTLLRDDDAGMLKGWSW 634
 QY 646 ---LSEA-----THSMALTDPYSGAFCP 666
 Db 635 DQALSEALSVASQNLTCRSWGEHRRPFTHTPLATQPPAWAGLLNP 679

Search completed: January 30, 2004, 13:18:32
 Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.
DM protein - protein search, using sw model
Run on: January 30, 2004, 13:10:12 ; Search time 46 Seconds
(without alignments)
6748.637 Million cell updates/sec

Title: US-09-769-699-2
Perfect score: 6294
Sequence: 1 MENTQKTVTPGVLGYVA.....DELFDLSGIPKIGNITWEM 1203
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525
Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_prodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4627	73.5	1194	Q9E1Y7	Q9E1Y7 cercopithec
2	3583	56.9	1208	Q39273	Q39273 equine herp
3	3371.5	53.6	1203	Q89549	Q89549 bovine herp
4	3327.5	52.9	1177	Q32611	Q32611 pseudorabie
5	3012	47.9	1197	Q89101	Q89101 herpes simp
6	2796.5	44.4	1190	Q92522	Q92522 turkey herp
7	2793.5	44.4	1190	Q92522	Q92522 turkey herp
8	2788	44.3	1191	Q92522	Q92522 turkey herp
9	2781.5	44.2	1190	Q92522	Q92522 turkey herp
10	2764.5	43.9	1190	Q92522	Q92522 turkey herp
11	2710	43.1	1191	Q92522	Q92522 turkey herp
12	1445	23.0	999	Q92522	Q92522 turkey herp
13	683.5	10.9	1132	Q40913	Q40913 kaposi's sa
14	675	10.7	1133	Q89904	Q89904 kaposi's sa
15	652.5	10.4	1103	Q92226	Q92226 murid herpe
16	651.5	10.4	1103	Q41928	Q41928 murid herpe

17	651.5	10.4	1127	12	Q93K9	Q93K9 callitrichi
18	645	10.2	1125	12	Q8B414	Q8B414 porcine lym
19	644	10.2	1125	12	Q8B408	Q8B408 porcine lym
20	633	10.1	1128	12	Q9YTQ7	Q9YTQ7 ateline her
21	623.5	9.9	1133	12	Q8UZD2	Q8UZD2 cercopithic
22	620	9.9	1132	12	Q9WRU1	Q9WRU1 macaca mula
23	618	9.8	1145	12	Q66611	Q66611 equine herp
24	596	9.5	1127	12	Q33360	Q33360 alcelaphine
25	593.5	9.4	1126	12	Q8JYD5	Q8JYD5 porcine lym
26	576.5	9.2	1142	12	Q8B424	Q8B424 porcine lym
27	576.5	9.2	1142	12	Q8B3X5	Q8B3X5 porcine lym
28	561.5	8.9	1134	12	Q99D22	Q99D22 bovine herp
29	505	8.0	1223	12	Q8OS31	Q8OS31 chimpanzee
30	484	7.7	1131	12	O56282	O56282 human herpe
31	401	6.4	1281	12	Q85425	Q85425 rat cytomeg
32	237.5	3.8	483	12	Q82172	Q82172 leporid her
33	235	3.7	84	12	O56864	O56864 feline herp
34	156.5	2.5	1841	2	Q9FE25	Q9FE25 streptomyce
35	143.5	2.3	178	12	Q64908	Q64908 alcelaphine
36	138	2.2	1869	2	Q8L334	Q8L334 aphanizomen
37	137	2.2	1504	2	Q9ZGA6	Q9ZGA6 streptomyce
38	129.5	2.1	919	5	Q9YA02	Q9YA02 drosophila
39	128.5	2.0	919	5	Q8IGV6	Q8IGV6 drosophila
40	123.5	2.0	1287	16	Q8EV67	Q8EV67 mycoplasma
41	120.5	1.9	788	16	Q8U9K3	Q8U9K3 agrobacteri
42	120.5	1.9	968	4	Q13025	Q13025 homo sapien
43	120.5	1.9	1555	5	Q8TIQ4	Q8TIQ4 trypanosoma
44	120	1.9	2152	16	Q9EX54	Q9EX54 streptomyce
45	119.5	1.9	701	16	Q92EA7	Q92EA7 listeria in

ALIGNMENTS

RESULT 1

Q9E1Y7 PRELIMINARY; PRT; 1194 AA.
AC Q9E1Y7
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SBDNA binding protein.
OS Cercopithecine herpesvirus 7.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=35245;
RN [1]
RP SEQUENCE FROM N.A.
RA Gray W.L., Starnes H.B., White M.W., Ashburn C.V., Mahalingam R.;
RT "Complete Sequence of the Simian Varicella Virus Genome."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ Databases.
DR EMBL; AF275348; AAG27202.1;
DR InterPro; IPR000635; Viral_DNA_bind.
DR Pfam; PF00747; viral_DNA_bd_1.
SQ SEQUENCE 1194 AA; 131968 MW; EBA7F3C841965897 CRC64;

Query Match 73.5%; Score 4627; DB 12; Length 1194;
Best Local Similarity 71.8%; Pred. No. 0;
Matches 867; Conservative 143; Mismatches 178; Indels 20; Gaps 8;

Qy	1	MENTQKTVTPGVLGYVACRVEDLDLEISPLAARSTDSLDLPLMRNLTVKFTTS	60
Db	1	MEATQKTVTPGVLGYVACRVEDLDLEISPLAARSTDSLDLPLMRNLTVKFTTS	60
Qy	61	SLAVVSGARTTGLAGAGITKLTTSHTFYPFVFGGKRVLPSSAPNLTACNAAREFF	120
Db	61	SLAVVSGRTTKTGLAGAGITKLTTSHTFYPFVFGGKRVLPSSAPNLTACNAAREFF	120
Qy	121	GFRCQCPFDGAVETGAEICTRGLGEPENITLYVTALKEAVFMCNVLHYGGDI	180
Db	121	GFNKRGSPPVDGAEITGADICNRIHLNPEHAFLYLIVTSLFKEAYMCSFLYYGIDN	180
Qy	181	VHINHGVDIRIPLPFFVQLEFMPDVNRLVPDPFNTHRSIGEGFVYPTFFYNTGLCHLIHDC	240

09/769,699 22-04
Search Notes

Db 181 VRIGRAVTRIPFLPIHVFDPVNVRLVTDNFNATORSIGENFYPTPLFNGNINCLLYDC 240
 Qy 241 VIAPMVALVRNVTA VARGAAHLAPDENHEGAVLPDITVYFQSSSGSTTTARGAREN 300
 Db 241 VISPIAVSLIRNVAVARGAAHLAPDENHEGAVLPDITVYFQSSSGSVNPRLPON 300
 Qy 301 DVNSTKSPSGGFERELASIMAAADTALHAENVFNTOIYEETPTDIKEWPMFIMGEGTLP 360
 Db 301 ----TSKSSQT-GFERELASIMAAADTALHAENVFNTOIYEETPTDIKEWPMFIMGEGTLP 360
 Qy 361 RLNALGSYTVARVAGVIGAMVFSNSALYLTVEDDSGMTAKOGGPGPSNRFVQFAGPHL 420
 Db 356 RLNLGSYTVARVAGVIGAMVFSNSALYLTVEDDSGMTAKOGGPGPSNRFVQFAGPHL 415
 Qy 421 AANPQTDREDDHVLSSQSGSNTSEFFSYDYLALICGFGAPILARLLFYLERCDAGFTGGH 480
 Db 416 AANPQVDRDGCWVFGSGTGVNTEFNVYDILALVCGSPQLLALLFYLERCDAGFTGGH 475
 Qy 481 GDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRPFGQATROPQVFGVTMNSQ 540
 Db 476 SDALQYVMFNSDIPCSLCEKHTRHLCVHTTQRLKQRPFGQATROPQVFGVTMNSQ 535
 Qy 541 YSCDPLGNVAPYLILKPGDQTEAAKATWQDTRATLERLFDLQERLLDRGAPCSSE 600
 Db 536 YSCDPLGNVAPYLILKPGDQTEAAKATWQDTRATLERLFDLQERLLDRGAPCSSE 595
 Qy 601 GLSSVVDHPTFRILDLTARIEQTTQPMKVLVETRDYKIREGLSEATHSWALTFDYP 660
 Db 596 GLNSTIIDHATFRALVFPKTRVQIQTEQIKVLVETRDYKIREGLSEATHSWALTFDYP 655
 Qy 661 SGAFCPITNELVXKTHLAVQDLALSCHQVYQVQVEGNGFNQPVLRFRFVLENG 720
 Db 656 SGSTCPITVFLVXKTHLAVQDLALSCHQVYQVQVEGNGFNQPVLRFRFVLENG 715
 Qy 721 GFSTRISITVLSEGPVSAPNPLTGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVFSGN 780
 Db 716 GFLSTRPVTVLPEGPPIAIDPTLQCAPA-CFDFGDLTKVSDVDIRDLVKNRVFSGS 774
 Qy 781 CTNLSEARARLVGLASAYORQKRVDMLGALGFLKQFHGLLFRGMPNPSKSNPQW 840
 Db 775 CTNLSEARARLVGLASAYORQKRVDMLGALGFLKQFHGLLFRGMPNPSKSNPQW 834
 Qy 841 FWTLLQONPADKLTREHETIAAVKRFTEEVAANFINLPTCIGELAQFVWMLILK 900
 Db 835 FWALLQONPADNLTSEEINTAIAKRFTEEVEANFINIAPTICIGELAQFVWMLILK 894
 Qy 901 YCHSQYLINTLSITGARPRDPSSVLHWIKOVTSAADITQAKALLEKTENL-PEL 959
 Db 895 YCHSQYFINTLTAVITGSKRPNPSSALHWIDKEITPTDLETHARLLIRNTEQMSPTM 954
 Qy 960 WTTAFTSTHLVRAAMNCRPMVLIGISISKYHGAAGNVRVFAQNWGLNGKNCVCLPTFF 1019
 Db 955 WISSFNSTNLVRAAMNCRPMVLIGISISKYHGAAGNVRVFAQNWGLNGKNCVCLPTFF 1014
 Qy 1020 DTRRRFIACPRGFCIFVTPGSSGNRETTLSQVRGIIIVSGGAMVQLAIYATVRAVGA 1079
 Db 1015 DTRRRFIACPRGFCIFVTPGSSGNRETTLSQVRGIIIVSGGAMVQLAIYATVRAVGA 1074
 Qy 1080 RAQMAFDDMLSLTDDFLARDLIELHDOIIQTLFTPTWVEGALVAIL-DEKTTAGG 1138
 Db 1075 RAEHMGFDDMLSLTDDFLARDLIELHDOIIQTLFTPTWVEGALVAIL-DEKTTAGG 1134
 Qy 1139 ETPNTLAFNPDSCPSHDTTNSVNLISGSISSGTVP---GLKRPDEDDLELSGIPK 1195
 Db 1135 SGCGNIAFNFTDNCDS-----NTEENAFPSYQLCSGVKPRDDDLLFDWSSIPOK 1186
 Qy 1196 HGNITMFW 1203
 Db 1187 M-SLTMDM 1193

RESULT 2

O39273
 ID O39273 PRELIMINARY; PRT: 1208 AA.
 AC O39273;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Counterpart of HSV-1 gene UL29 and VZV gene 29.
 GN 31.
 OS Equine herpesvirus 4.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirus.
 OX NCBI_TaxID=10331;
 RN [1] TaxID=10331;
 RP SEQUENCE FROM N.A.
 RC STRAIN=NS80567;
 RX MEDLINE=98264497; PubMed=9603335;
 RA Telford E.A., Watson M.S., Perry J., Cullinane A.A., Davison A.J.;
 RT "The DNA sequence of equine herpesvirus-4";
 RL J. Gen. Virol. 79:1197-1203 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NS80567;
 RX Telford E.A., Watson M.S., Perry J., Cullinane A.A., Davison A.J.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF030027; AAC59547.1;
 DR InterPro: IPR000635; Viral_DNA_bind.
 DR Pfam: PF00747; viral_DNA_HF; 1.
 SQ SEQUENCE 1208 AA; 130607 MW; 17699FED3238C4CB CRC64;
 Query Match 56.94; Score 3583; DB 12; Length 1208;
 Best Local Similarity 56.44; Pred. No. 2,7e-286;
 Matches 689; Conservative 194; Mismatches 305; Indels 34; Gaps 13;
 Qy 1 MENTQKTVVPTGPGVYACRVEDDLLEISFLAARSTDSLALLPLMNLVETFTS 60
 Db 1 MESAPKTVSLPVSGLGVYAIQNTFMETALTMAKSIDSLAVLVIRGLVYQTFIT 60
 Qy 61 SLAVVSGARTTGLAGAGITLKLTTSHFVPSVFPVPHGKHLVPSAAPNLTRACNAARERF 120
 Db 61 NVAVVAGSKITGLGAGITLKLTPSHPTPNAFYVGSVFGASKAPNLTRACELARRP 120
 Qy 121 GFSCQCPVVDGAVETTGABICTRLGLEPENTILYLVVLTALFKAEVPMCNVFLHYGGLDI 180
 Db 121 GFSPFSPVVDNAVETSGEBCASLNLSPETTLVLTVTETKEMVMCMFTLHYGTST 180
 Qy 181 VTHNGDVIRLPLFPVQLFMPDVNRLVPDPNTHRSIGSGFVYPTPYNTGLCHLHDC 240
 Db 181 VTTHQGAVKIPIIPVQLYMPDVNRLAAEPFNKHSIGDEFFVSKPFFNSDLCLLLHY 240
 Qy 241 VIAPMVALVRNVTA VARGAAHLAPDENHEGAVLPDITVYFQSS--SSGTTTARGAR 298
 Db 241 VLGPAAVALVRNLDGVARGAAHLALDENHEGAVLPDITVYFQSSGSGRGTOR 300
 Qy 299 RNDVNSTKSPSGGFERELASIMAAADTALHAENVFNTOIYEETPTDIKEWPMFIMGEGT 358
 Db 301 QGD-GSGLEKNGSSGIERELASIMAAADTALSVDSIMGAGVYDTLPSVEDLPI-LSVGDD 358
 Qy 359 LPRNLALGSYTVARVAGVIGAMVFSNSALYLTVEDDSGMTAKOGGPGPSNRFVQFAGP 418
 Db 359 RERLEALGAYASRLSGLNGAMVFSANSVLVWTEVDDOGPADGKAS-NPSYHLPYLIAP 417
 Qy 419 HLAAMPQTDREDDHVLSSQSGSNTSEFFSYDYLALICGFGAPILARLLFYLERCD 472
 Db 418 YVAGNPQTDKDRVLOHTAQPAPAPINGSNQSFSLYALACGFCPOLARILFYLERCD 477
 Qy 473 AGAFTG-CHGDALAKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRPFGQATROP 531
 Db 478 AGTFGRNETDALRYLANTLIESEVPCGLCTPATRPACAHHTLHRLRQRPFGTPTVAPI 537
 Qy 532 GYFGTMSQVSDCDPLGNVAPYLILKPGDQTEAAKATWQDTRATLERLFDLQERLL 591
 Db 538 GYFGTMSQVSDCDPLGNVAPYLILKPGDQTEAAKATWQDTRATLERLFDLQERLL 596

```
Y 592 DRGA-----PCSEGLSSVVDHPTFRILDLRLARIEQTTPQMKVLVETRDYKIREGL 646
b 597 DEALAHAGTCSAS--TGVRKDQASFINLLSTIKITEGAEEQFVRTLVEVRDFKIREGL 654
Y 647 SEATHSMALTEPDYSGAFCPITNPLVFKTHLAVQDLALSOCHCVFYGOQVEGFRNNOF 706
b 655 ADANTHMSISLDPYSSPCPVSFLSRRTIFAVLDQVLVSQCHCLFYGOQVEGFRNNOF 714
Y 707 QVLRRRFDVFNPGFISRTSITVTLSEGPSVAPNPTLQGDAPAGRTPDGDLARVSVEVI 766
b 715 QVLRRLRFLDMLNGGFITAKTVTVTSVSGVTAPNLTLPSPSEPTTKDYDGMARVSMEVL 774
Y 767 RDIRVKRWVSGCNTLSEARARLVGLASQYQKRXVDMHLGALGFLIKQPHGLFP 826
b 775 RLRLIKRWLVSGNANSEARARVAGWASAYRRPERGSMILNGAVGLVQPHKVLFP 834
Y 827 RGMPPNSKPNPQFWTLQRNQMPADKLTHEEITIAAVKRFTEEYAAINFINLPPTCI 886
b 835 RGHPPGIDTPNPQFWTLQRNQMPARLLSKEDIEITAIKFSHSYSAINFINLTPNI 894
Y 887 GELAQFYMANLILKYCDHSOYLINTLSITIGARRPPSPSVLHWIRKDVTSAADITQOA 946
b 895 GELAQFYANLVLYKCDHSQVFINGLTAIVVGSRRPRDPAALAWINRTINGASDVPEAA 954
Y 947 KALLEKTENLPELWTTAFTSHLVRAAMNQRMVVLGISIKYHGAAGNNRVFOAGNWSG 1006
b 955 QEVLOQLGSGNPAWGTGTASTNMVYVWDQPMVVGISIKYSGSAGNNRVFOAGNWSG 1014
Y 1007 LGGKQNVCLFTFTRFIIACPRGGFCIPVTGSSGNRTTLLSDQVRGIIVSGGAMVQ 1066
b 1015 LGGKQNVCLMAFDRTRFVLACPRVGTCEAGFGMGARENLTSEQIRSVSDGGMVQ 1074
Y 1067 LAIVATVRAVGARQAQWAFDDWLSLTDDEFLARDLELHQIICLTETPWTVEGALEAV 1126
b 1075 TAVFSVLTALGARTQHLAVDDWIGLVDDEFLASLDNALNAVVDQF-GEWSVEAAQDMI 1133
Y 1127 KILDEKT-----TAGDGETPTNLAFNEDSCFPHDITTSNVLNISGNSITVPGKLRPP 1181
b 1134 RTMDAQTNMGVVSTGDG-----AFDFGACVDANQSSTTFNMGPA--SSSAPAGQKRFH 1185
Y 1182 EDELEFDSIGPIKSGNTMEM 1203
b 1186 PDDILFDNGAPPEKXSLTFDM 1207

RESULT 3
D Q89549 PRELIMINARY; PRT; 1203 AA.
T C Q89549;
T 01-NOV-1996 (TReMBLrel. 01, Created)
T 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
T 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
E UL29.
N Bovine herpesvirus 1.
S Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
C Alphaherpesvirinae; Varicellovirus.
X NCBI TaxId=10320;
X [1]_
P SEQUENCE FROM N.A.
C STRAIN=COOPER;
A Schwyzer M., Vlcek C., Lowery D.E., Bello L.J., Meyer G., Misra V.,
L Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
P [2]
P SEQUENCE FROM N.A.
C STRAIN=COOPER;
A Meyer G., Vlcek C., Paces V., Pastoret P., Thirty E., Schwyzer M.;
L Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
P [3]
P SEQUENCE FROM N.A.
C STRAIN=JURA;
A Schwyzer M., Vlcek C., Lowery D.E., Bello L.J., Meyer G., Misra V.;
```

```
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=JURA;
RA Schwyzer M.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; X98205; CAB01596.1;
DR EMBL; X94677; CAAG4336.1;
DR EMBL; AJ004801; CAAG0104.1;
DR InterPro; IPR000635; Viral DNA bind.
DR Pfam; PF00747; viral DNA bp. 1.
SQ SEQUENCE 1203 AA: 127409 MW: 8299064566A9654F CRC64;

Query Match 53.6%; Score 3371.5; DB 12; Length 1203;
Best Local Similarity 53.6%; Pred. No. 8.1e-269;
Matches 652; Conservative 197; Mismatches 340; Indels 27; Gaps 16;

QY 1 MENTQKTVVPTGPIGVYVACRVEDLDEEIEFLAARSTDSLDALLPLMRNLTVKETS 60
Db 1 MDAATAKTVALAPGAGFYVCDAAARDLDKALLAARSDSLAVPLVRLGVETAPAP 60
QY 61 SLAVSGARTGLAGAGITLKLTTSHFVPSVVFHGGKHLVLPSSAAPNLTRACNARRP 120
Db 61 NYAVAGTAKTGLGAGLTAKLTPSHYPNVVFHGGKHLVLPSSAAPNLTRACNARRP 120
QY 121 GFSRCQPPVDCGAVETGABICTRLGLSPENTILVLTALFKEAVFNCVFLHYGGDI 180
Db 121 GFSAFAGAPVDGAVETTAEDICRAVGASPEVALVLAITTEAPKEIVYCNITFLHYGGA 180
QY 181 VHINQDVIRIPLFPVQLPMDVNLVDPDPNTHRSIGEGFYPTFPYNTGLCHLIHDC 240
Db 181 VQVGAEAVRVLFPVQLYMPDNRVNLPEPNARQALGEOLAVRPFYNAALCELLGY 240
QY 241 VTAPNAVALRVNTAVAGAAHLAPDENHGNLPPDITVITYFOSSSGTITARGARN 300
Db 241 VLGPAAVALRVNTLDAVAGAAHLAFDESHGALPPDPVCFVTF-DQAGRGSGSGARG 299
QY 301 DVNSTSKPSPGSGFERRLASIMAAATLHAELVIFMTGIVEETPTDIKEWPFIFMEGL 360
Db 300 APGAKAAPAGGVERRLASVMAADTVSIEANSTSVFDEEDVACVDDWPMLOGAADA 359
QY 361 RNALGSYTVARVAGVIGAMVFPNSALYLTVEBESGNTAKDGGPSPFNRFYOPAGPHL 420
Db 360 KLDALGAVYVRLAGLVGAMVFPSSNSVLHMTVDGGAADAKDGA-AAAGHFHYQTAAFY 418
QY 421 AANPOTDRDGHVLSOSTG-----SSNTEPSVDYVALICGGAFLARLLFYLERCDAG 474
Db 419 AGNPRCDKDKPELPOTGAGPAVINGAQEAFALDHIALACGCTPOLLARMLFYLERCDAG 478
QY 475 AFTGGHG-DALKYVTGTFTDSEIPCSLCEKHTRPVCATTVHRLRQMRPFQOATROPIGV 533
Db 479 AFAGRNDMDALYVAVASTLEGDVFCGLGSRDDRHACATTLHLRHLRFPFGAPTRSP 538
QY 534 FGTMNSQISDCCPLGNYPYLILRKPGQCTEAAKATQDTYRATLERLFIIDLEQLLDR 593
Db 539 FGTMNSAYSDCCVLGNYSYLSALRPG-ADENASIMQATYRAAVENVALDQOORLLAH 597
QY 594 GAPCSSEGLSSVIVDHPTRRILDLRLARIEQTTPQMKVLVETRDYKIREGLSEATHSM 653
Db 598 DAQSAQAQ-LERAITDHSFSGALAAIQNTVQAIEAFVRLGVEDDFQREALYEAHTL 656
QY 654 ALTFDYPYGAFCPIITNPLVFKTHLAVQDLALSOCHCVFYGOQVEGFRNNOFVLR 713
Db 657 SLALDPYSTAVCPATAFLFRSRVLAQDLALSOCHGIFCQGPVDGFRFRAQFQVLR 716
QY 714 FVDLNGGFGISTRTSVTLSEGPSVAPNPTLQGDAPAGRTFDGLARVSVIRDIRVKN 773
Db 717 FMDLLNGGFLITRTVTVTLAEAAVAPNLAAQTEPPARDMDGDISKVSLEVFKEMRVKN 776
QY 774 RVYFSGNCTNLSEARARLVGLASAYQKQK-RVMDLHGALGFLIKQPHGLLFRPGMPPN 832
Db 777 RVWFSGNANMSEARARLVGLAGAYQKPSGGVNLSSPLGLVFKQPHKLFNGKPPG 836
```

QY 833 SKSPNPFQWTLTQONQPADKLTHERBITTAAVQRTEYAAINFINLPTTCIGELAQF 892
 DB 837 SPTNPQWFWTLTQONQPADKLTHERBITTAAVQRTEYAAINFINLPTTCIGELAQF 896
 QY 893 YMANLILKYCDHSOYLINTLTSIITGARRPRDPSPVHLWIRKDVTSAADETQAKALLEK 952
 DB 897 YMANLILKYCDHSOYLINTLTSIITGARRPRDPSPVHLWIRKDVTSAADETQAKALLEK 956
 QY 953 TENIPELWTTAFTTHLVRAAMNORPMVLGISISKTHGAGNRRVFOAGNWSGLNGGKN 1012
 DB 957 ABAEPYVWAGTFAASNLVRSYMATPRAVVLGLSISKYNGSAGNRRVFOAGNWSGLNGGKN 1016
 QY 1013 VCPLEFTRTRRFFIACPRGFCFVTPGSSGNRETTLSDOVRGII--VSGAMVOLAIYA 1071
 DB 1017 VCPLEFTRTRRFFIACPRGFCFVTPGSSGNRETTLSDOVRGII--VSGAMVOLAIYA 1076
 QY 1072 TVRAVAGARQAHMADFDDWLSLTDDEFLARLDELHDOIIQTLETP--WTVEGALEAVKIL 1129
 DB 1077 AVLQALGPRVHEMDLDDWAALVEDEFFAQSMVELTE---RAAARPGWSPGGAADMLREL 1133
 QY 1130 DEKTTAGDGETPTNLAFNEDSCPSHDTTNSVLN--ISGSNISGSIVPGIKRPPEDELEF 1187
 DB 1134 ELEBAAEAE--PAGAFDFGACASAPADATYAFGGPVAG---AGPAAPGAKR--PDLELEF 1187
 QY 1188 DLSGIPKIKHGNITMEM 1203
 DB 1188 EM-GVPDKRALITDM 1202

RESULT 4
 O92611 PRELIMINARY; PRT; 1177 AA.
 ID AC O92611; 1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE DNA-binding protein.
 GN DB.
 OS Pseudorabies virus.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirus.
 OX NCBI_TaxID=10345;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TNL;
 RA MEDLINE=98455382; PubMed=9784061;
 RW S. L., Hsiang C.-Y., Ho T.-Y., Chang T.-J.;
 RT "Identification, expression, and characterization of the pseudorabies
 virus DNA-binding protein gene and gene product."
 RL Virus Res. 56:1-9(1998).
 DR EMBL; U80909; AAC63429.1; -
 DR InterPro; IPR000635; Viral_DNA_bind.
 DR Pfam; PF00747; viral_DNA_bp; 1.
 KW DNA-binding.
 SQ SEQUENCE 1177 AA; 125408 MW; BA87AF9CFC961707 CRC64;

Query Match 52.9%; Score 3327.5; DB:12; Length 1177;
 Best Local Similarity 53.5%; Pred. No. 3.4e-265;
 Matches 655; Conservative 182; Mismatches 318; Indels 69; Gaps 17;
 QY 1 MENTQKTVVTPGTPLGVYACRVEDLDLEISFLAARSTDSIALPLMNLTVKTFTS 60
 DB 1 MEAAKTVTVRAAPLGYYTPIALRRDLALLVARSADDAAVAPLVGLTVFAGFAG 60
 QY 61 SLAVVGARTTGLAGAGITLKTSTHTPVPVFGGKHYLPSSAAPNLTRAANARERF 120
 DB 61 HVAVVAGARTTGL--GGGLTLKLPNHPNVPFFHNGDCVPPSSAAPALSRACEAARAF 119
 QY 121 GFSCQGFVVDGAVETTGATCIRGLGEPENTILYLWTFALPKVPMCNVFLHGLDI 180
 DB 120 GFSAYR--TPVDNAEETTGAVRCGLAADAHAAYLVADGPKFVAVLHGLHGGAGT 178
 QY 181 VHINHGDIVIRIPLFPVQWFLPVDVNLVDPDFTNTHRSIGBGFVYPTFYNTGLCHLHDC 240

RESULT 5

DB 179 VSINGHEAWRVLPVPLFMDVNLVADPNFNAKRSISEEFYPRFFNGPLCRLHGY 238
 QY 241 VIAPMAVALRVNTAVARGAAHLAFDENHGAVALPDDITVYFOSSSSGTTTARGARN 300
 DB 239 VLGPAAVATRVNLDNARGAAHLAFDENHSAVLPADVTFTLFEQRRRGDDAA----- 292
 QY 301 DYNSTSKPSPGGFERRLIASIMAAATLHAHVINTGIIYEETPDIDKEMPMF----- 352
 DB 293 -----AGGLERBMASVMSADAALSLEALVAAGVDEEPPALDDWVLEAGAKDG 342
 QY 353 ---IGMEGTLPRLNALGSYVARVAGVIGAVWSPNSALVLTVEDESGMTEAKDGGPGPS 408
 DB 343 GAAAASAGAPVQAAALGAVVSRRAAGLVGALVFNSSNSVLYLTVDDGAADAKKEGAGPS 402
 QY 409 FNRFYQFAGPHLAANPOTDRDGHYL---SSQSTGSSNTFESVDYLALICGFGAPLIAL 464
 DB 403 FNRFYQVAAAPYLAGNPOTDKDGRVLLHTASOPATAPGNHDFAMHDLVWAGFCPOLIARV 462
 QY 465 LFYLERCDAGPTG--GHGDALKVYGTGTFDSSEIPLCSICEKTRPVCAHTTVHRLRQMPRE 533
 DB 463 LFYLERCDAGPTG--GHGDALKVYGTGTFDSSEIPLCSICEKTRPVCAHTTVHRLRQMPRE 531
 QY 524 GQATROPITGVFTWNSOYSDCDPLGNYPALILIRKPGDQTEAAKATMQDTYRATLERLEI 583
 DB 522 SARRESPMGVFTWNSAYSDCDVLGNVASYGALKRPND--SEPPKALMQDTYRAAVDRLLA 580
 QY 584 DLEQERLLDRGAPCSSSEGLSSVVDHPTFRRIIDTLRARIEOTTQPMKVLVETROYKIR 643
 DB 581 DVAGAR-----IGETVDEHAGFRHALRALDVTVEQAAARFVRLVETROFKIR 628
 QY 644 EGLSEATHSMALTEDPYSGAFCPITNPLVKRTHLAVVQDLALSCQCHCVFYGQOQVEGNFR 703
 DB 629 DALVDANHTWLSLDIPSGALCPATSLARLTLLAVLODLALSCQCHGVHGOVVEGNFR 688
 QY 704 NQOPVLRRPVDLFGNGFISTRTITVTLSEGVSAVNPFLGODAPAGRTFDGDLARVSU 763
 DB 689 NQOPVLRRPVDLFGNGFISTRTITVTLSEGVSAVNPFLGODAPAGRTFDGDLARVSU 747
 QY 764 EVIRDIRVQRVVFSGNCTNLSEAAARLVGLASAYORQEKRVDMHLGALGFLKOFHGL 823
 DB 748 EVIRDIRVQRVVFSGNCTNLSEAAARLVGLASAYORQEKRVDMHLGALGFLKOFHGL 807
 QY 824 LFPKMPNPKSPNPQWFWTLTQONQPADKLTHERBITTAAVQRTEYAAINFINLPTTC 883
 DB 808 LFPKMPNPKSPNPQWFWTLTQONQPADKLTHERBITTAAVQRTEYAAINFINLPTTC 867
 QY 884 TCIGELAQFYMANLILKYCDHSOYLINTLTSIITGARRPRDPSPVHLWIRKDVTSAADE 943
 DB 868 GTVAELAQFYMANLILKYCDHSOYLINTLTSIITGARRPRDPSPVHLWIRKDVTSAADE 927
 QY 944 TQAKALLEKTENIPELWTTAFTTHLVRAAMNORPMVLGISISKTHGAGNRRVFOAGN 1003
 DB 928 RAAREVLDAPRD--DTWVATYTSHELLRSVWASRPLVWLGLGVSKYHGWAGNRRVFOAGN 985
 QY 1004 WSGLNGKNVCPLFTFDRTRRFFIACPRGFCFVTPGSSGNRETTLSDOVRGIIIVSGGA 1063
 DB 986 WSGLNGKNVCPLFTFDRTRRFFIACPRGFCFVTPGSSGNRETTLSDOVRGIIIVSGGA 1045
 QY 1064 MVQLAIYATVAVRARGAQAQHMADDDMLSLTDDFLARDLEELHDOIIQTLETPWTVEGAL 1123
 DB 1046 LVQTVAVVAUHALGARTQHLBEDDWRIVDDFLAALAEINGRVADR--DGRWSVEAAA 1104
 QY 1124 EAVKILDEKTTAGDGETPTNLAFNEDSCPSHDTTNSVLNISGSNIS--GSTVP--GLKR 1179
 DB 1105 ELVDLEQGTGADGGGRET--AFDFGACGAGGD-----AGAGLAPASLAPAEILGGR 1154
 QY 1180 PPEDDELFDLSGLPIKHNITMEM 1203
 DB 1155 PPEDDELFDLSGLPIKHNITMEM 1176

```

69101
D Q69101 PRELIMINARY; PRT; 1197 AA.
C Q69101;
T 01-NOV-1996 (TrEMBLrel. 01, Created)
T 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
T 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
E DNA binding protein ICp8.
S Herpes simplex virus (type 2).
S Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
C Alphaherpesvirinae; Simplexvirus.
X NCBI_TaxID=10310;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=KN;
X MEDLINE=93228441; PubMed=8385914;
X Toh Y., Tanaka S., Iau Y., Mori R.;
T "Nucleotide sequence of the major DNA binding protein of herpes
T simplex virus type 2 and comparison with the type 1 counterpart.";
L Arch. Virol. 129:183-196(1993).
L R EMBL; D10658; BAA01507.1;
R InterPro; IPR000635; Viral DNA bind.
R Pfam; PF00747; viral_DNA_b.1.
R Q SEQUENCE 1197 AA; 128470 MW; AA3ADA75B8865BFE CRC64;

Query Match 47.9%; Score 3012; DB 12; Length 1197;
Best Local Similarity 49.0%; Pred. No. 4e-239;
Matches 595; Conservative 212; Mismatches 368; Indels 40; Gaps 16;

Y 4 TQKTVTVTGLGVY--ACRVEDLDLSEIFLAARSTDSALLPLMRNLTVKTTSS 61
b 7 TTTTVKVPVGMVYVGRACPAEGL--LSLLSARSQDADVAVAPLIVGLTVESGFAN 64
Y 62 LAVVSGARTTGLAGATILKLTTHSFVPSVVFVGGKHVLPSSAAPNLTRACNAARERFG 121
b 65 VAAVVGSTTGLGTVASVSLKMPHSYVSVVFGGRLHAPSTQAPNLTLCEARARHFG 124
Y 122 FSRQGGPVDDGAVETTGAEITRIGLEPENTILYLVVTLFKEAVFVNCVPLVGGLDIV 181
b 125 FSDVAPRPCDLKHETTGDCALCERLGLDPRALLYLIVITEGPREAVCISNTPLHLGMDKV 184
Y 182 HINQGVIRIPLEVPQMPDVRNLRVPPNTHRSIGEGFVYPTPVNTGLCHLHDCV 241
b 185 TIGDAEVHRIPIVPLQMPDPSRVIAIDPFNCNRSIGENYLPFPNPLARLLFEAV 244
Y 242 IAPMAVALRVNVTAVARGAAHLAFDENHGAFLPPDITYTFQSSSGTTTARGARND 301
b 245 VGPAVALRVNVDVAVARAAHLAFDENHGAALPADITTAFAESOG--KPGRGAR--- 299
Y 302 VNSTSKPSGPFERLASIMAAHTALHAELVFTGTYEETPTDIKEWPMFIMEGTLP 361
b 300 --DAGNKGPFAGGFEQRLASVWAGDAALALEISVMAVDFEPPDITWPLLEGQETPAAR 357
Y 362 LNALGSYTVAVGVIGAMVFPNSALYLTVEVDSGMTAKDGGPSPFNRFYQFAPGPHLA 421
b 358 AGAVGAYLARAGLVAMVFTNSALHTEVDDAGPADPKDHSK-PSFYRFFLVPGTHVA 416
Y 422 ANQTDTRDGHVL-----SSQSTGSSNTEFSVDYALICGFGAPLARIALLFYLERCDAG 474
b 417 ANPQLDREGHVVPVGGREPTAPLVGGTQ-EPAGEHMLAMCGFSPALLAKMLFYLERCDGG 475
Y 475 AFTGCHG-DALKYVVTGTFDSIPCSLCEKHTFPCVANTVHRLRQRPQATQPIGV 533
b 476 VIVGQEMDMVFRYVADSGQTDPVNCNLTCTFTRHCAHTTLMLRARPFPASARGAIGV 535
Y 534 FGTWNSQSDCDPLGNAYPIYLIRKPGQTEAAKATWQDTRYATLERLFLDLEQERLLDR 593
b 536 FGTWNSAYSDCDVLGNAYAFSAKX-ADGSENTRTIMQETVYRATERVMAELEALQYVDQ 594
Y 594 GAPCSSEGLSSVVDVHTPFRILDTLRARIEQTQTFQMKVLVETRDYKIREGUSEATHSM 653
b 595 AVPTALGLETIIGNREALHTVWNKIKOLVDREVEQLMRNLIEGRNFKFRDGLAEANHAM 654
Y 654 ALTFDPYSGAFCPITNFLVKRTHLAVVQDLALSQCHCVFQQVQGVQGVNFRNQFVLRRR 713

```

```

Db 655 SLSDPYTCGCPCLQLLARRNLAAYQDLASUQCHGVFAGQSVGEGRNFRNQFQVLRRR 714
Qy 714 FVDLFNGGFISTRSITVTLSEG-PVSAPNPTLQGDAPAGRTFDGDLARVSVEVIRDIRVK 772
Db 715 VMDLPNNGFSAKTLTVALSEGAACAPSLTAGQAPAESFEGDVANVTGLGFPKELAVK 774
Qy 773 NRVPVSGNCTNLSEARARLVGLASAYQOEKRVDMHLAGLGFLLKQFHGLLFFRGMPPN 832
Db 775 SYVLFPAGASANASEAKARVASLQSAQKPKRVDILLGLPLGLLKQFHAVIFPNKPPG 834
Qy 833 SKSPNPQFWTLLQRONMPADKLTHEEITTTAAVKRFTTEEYAAINFILNPLPTCIGELAQF 892
Db 835 SNQPNQFWTALQPNQLPARLLSREDIETAFIKRPSLDYGAINFILNPNVSELAMY 894
Qy 893 YMANILKYCDHSOYLINTLSITGARRPRDPSPVLHWIRKDVTSAAIDTQAKALLEK 952
Db 895 YMANQILRYCDHSTVFINTLTAVIAGSRSPSPVAAAAPQ---GGAGLEAGARALMDS 951
Qy 953 TENLPELWTTAFTSLVRAAMNQPMVVLGISIKYHGAAGNRRVFOAGNWSGLNGGKN 1012
Db 952 LDAHFGAWTFMFASCNLLRPVMAARPMVVLGISIKYHGAAGNDRVFOAGNWSLLGKN 1011
Qy 1013 VCPLETFDTRRFFIACPRGGFICPVTGPSSGNRETTLSDDQVGIIVSGAMVQALAIYAT 1072
Db 1012 ACPILLIFDTRKFLACPRAGFVCAASSLGGGAHEHSLCEQLRGLIIEGGAIVASSVFVA 1071
Qy 1073 VYRANGARAHVAFDDWLSLTDDEFLARDLSELDHDIITQLETPWTVGAL-----EAVKI 1128
Db 1072 TVKSLGPRTQQLQIEDWLALLEDEYLSEMMEFTRALERGHGEMSTDAALEVAHEAEL 1131
Qy 1129 LDEKTAGTAGEPTNLAFNFCSEPSHDTTNSVLNLSGNTSGSTVPGCLKRPPEDDELFD 1188
Db 1132 VSQLGAGE-----VNFQDGFQDDEDDHAASFGGLAAA-LAGAAGVARKAFHGDGDFG 1183
Qy 1189 LSGIPIKGNITWEM 1203
Db 1184 -EGPPEKK-DLTLD 1196

RESULT 6
Q9PZ52 PRELIMINARY; PRT; 1190 AA.
AC Q9PZ52;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Homolog of HSV-1 glycoprotein.
OS Turkey herpesvirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Marek's disease-like viruses.
OX NCBI_TaxID=10390;
RN [1]_TaxID=10390;
RP SEQUENCE FROM N.A.
RC STRAIN=HPR24;
RA Kato K., Jang H., Izumiya Y., Cai J., Teushima Y., Miyazawa T.,
RA Kai C., Mikami T.;
RT "Identification and Transcriptional Analysis of the Marek's Disease
RT Virus Serotype 2 Genes Homologous to the Glycoprotein B (UL27), the
RT ICPI8.5 (UL28) and the Major DNA-binding Protein (UL29) Genes of
RT Herpes Simplex Virus Type 1.";
RL J. Vet. Med. Sci. 0:0-0(1999).
DR EMBL; AB024711; BAA83753.1; -.
DR InterPro; IPR000635; Viral DNA bind.
DR Pfam; PF00747; viral_DNA_b.1.
SQ SEQUENCE 1190 AA; 130339 MW; D9291E2EEAA7F59D CRC64;

Query Match 44.4%; Score 2796.5; DB 12; Length 1190;
Best Local Similarity 45.3%; Pred. No. 2.6e-221;
Matches 546; Conservative 235; Mismatches 396; Indels 29; Gaps 13;

Qy 6 KTVTVPTGVLGVVACRVEDLDLSEIFLAARSTDSALLPLMRNLTVKTTSSLAUV 65

```

Db 5 KSVKITGGPTGYVYAAPTSSMPADLDSIFAAKSNDCSDAILPLVSLGTVLEADFMNVAUV 64

Qy 66 SGARTTGLAGAGITLKLTHSHFYPSVFVFGKHVLPSSAAPNLTRACNAARERFGFSRC 125

Db 65 AGTKTTGLSGSTTLKLVPHYHPCVVFHGGECIKPTKAPNLTRACDLARGFGYS-T 123

Qy 126 QGPPVDGAVETGAEICTRIGLEPENTILYLVVTALEKAVFNCVNFVFLHYGGDIYHINH 185

Db 124 YSPPVATSFETTGEOICDNLGMNPQETMLVLTVELFKEAVYLCNSLYHYGGTVGSING 183

Qy 186 GDVIRIPLFPVQLEPMPDNRLVPDPFNTHRSIGEGFVYPTFPYNTGLCHLHDCVIAPM 245

Db 184 VDVRRILYPLHLVFPDNVADPSTKRALGEGALMPKAFVNSLCLLHGYVISTA 243

Qy 246 AVALRVNVTAVAGAAHLAFDENHEGAVLPDITYTYFOSSSGTITARGARNVDNST 305

Db 244 AVGLRVNVDIAICAAHLSDENHEGTLFPADTAFAFNAFTPIESASKSOYKAGKEGMEL 303

Qy 306 SKPSPSGGFERRLIASIMAAATLHAIEVINTGIVETPTDIKWPMPFIMEGTLPRLNAL 365

Db 304 S----GGYERRTASLASDATLSIENVIAISVYESIPDVKNPIYCSPIGYTDRVAL 359

Qy 366 GSYTARVAGVIGAMVFPNSALYTEVEDSGMTEAKD--GGPGPSFNRFYQFAGPHLAAMP 424

Db 360 SAYMARVAGLVGAMVFPSSNSVIYVTEVGEASSADGKDTSTTAAAFYRFFQIAAPHLASNP 419

Qy 425 QTRDGHVLS-----SOSTGSSNTEFSVDYLALICGECAPILLARLLVFLRCDAAGTGGH 480

Db 420 LIRBDGKPIGEDLSKATSAPEISYDYLILACGFCPLLARFLYBERCDGGSQACH 479

Qy 481 G-DALKYVITGTFSSBPCSCIXEHTRPVCAHTVHRLRQMRPFQOATQPIGVFGTMS 539

Db 480 DLDTVKFVSSAMDADVPCDLCKASRIYCAHTTIKRLEYLPKFGYQMRGAMGLFGSMTN 539

Qy 540 QYSDCDPLGNAPVLIILKRGDQTEAATAWQDYATATLERLIDLEQERLLDRGAPCSS 599

Db 540 NYCDVNALGSYAPFSTLKR--SEGETSRSMQDYTKLTVERNKALEKXEGLLTCEPDADM 597

Qy 600 EGLSSVIVDHTFRRILDTLRARIEQTTCFMKVLVETRDYKIREGLSEATSMALTFDP 659

Db 598 TPADAILRDGKSFMRALSTWANIIESEAGQLMENLREYNIREGLGATHTLSLAIEP 657

Qy 660 YSAGFCPTNLFVRLTHLAVQDLALISQCHVYQGOQVGRNFRNQFQVLRARRFDLFN 719

Db 658 YSSGICFVLSFLSRRTIIIVQDMALSSQSMVHQVQVEARNFRTQFQVLRARRFDLQN 717

Qy 720 GGFSTRISITVTLSEGVSPAPNPTLQDAPAGRTFDGLARVSVEVIRDIRVKNRVFSG 779

Db 718 AGTITSKNIITVLEDQOISVPDPSKQHDPLAVHMEGDLVKVTFEIPREFKVNKVMFVG 777

Qy 780 NCTN-LSEAAARLVGLASAYQOEKRVMDLHGALGLFLKQFHGLLFFRGMPNPKSNP 838

Db 778 GVSSTVSDATKSLAGMIEAYQPAKAMVNLGFLGALKRYHTQLFPNVKMPNGTTPNA 837

Qy 839 QWFTWLLQRNQMDKLTHEIITIAVAKRTEIEYAAINFILPPTCIGELAQFYMANLI 898

Db 838 LWFWILLQRNQLPAGLSKDAEENTSFIRKFTNSYADMYNIISPTCFGELAQFYLIANT 897

Qy 899 LKYCDHSQYLINTLTSITIGARFPRPSSVLHWRKDVTSAAIDIEQAALKEKTENLPE 958

Db 898 LRYCSHKHPFNITISALVATSKPRDPAWVLPWIEHLTQGSVDVAPAAQQLKNICDHKE 957

Qy 959 LWTATSTHLVRAANQRMVVLGTSISKYHGAAGNVRVQAGNWSGLNGKNVCPLFT 1018

Db 958 AWCAAFSSNTLVGPIVASKPFPVIAVYSISKYHGMAGSTKVFQSGNMGNTMGRCVIMS 1017

Qy 1019 FDRTRFIACPRGGFCPTVGFSSGNREFTLSDQVRGIIVSGAMVQALIAIVTVRAG 1078

Db 1018 FNRTHFVACPVAGFVBSQTFGSSGKLTLLVDRAVLSSECGAPHAATVIALKVG 1077

Qy 1079 ARAOHAFDMLSLTDDFLARDLELHDQIIQTLETPTWTEGALEAVKILDEKTTAGDG 1138

Db 1078 DRVROMELDDWMEITNDAYTSSILDEINKQ-VEGCEGWSVDAATAILAKEMVDMKALPL 1136

Qy 1139 EPTTNLAENFDCBPSHDTTSNVNLISGNSISGTVPLGLKRPDEDDLFOLSGIPKHG- 1197

Db 1137 DGPT--FNYDALDENTERRAD-----GPSI--LEFLAKRQSD--VFDLEVPVEKRAP 1183

Qy 1198 NITMEN 1203

Db 1184 GLSVDN 1189

RESULT 7

Q9QTB9 PRELIMINARY; PRT; 1190 AA

AC Q9QTB9

DT 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)

DE UL29 protein

GN ORF 37 OR UL29

OS Marek's disease virus serotype 2 MDV2, and

OS Gallid herpesvirus 3

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Alphaherpesvirinae; Marek's disease-like viruses.

OX NCBI TaxID=36353, 35250;

RN (1)

RP SEQUENCE FROM N.A.

RC SPECIES=Gallid herpesvirus 1 (serotype 2); STRAIN=HPRS24;

RA Jang H., Cai J., Izumiya Y., Murakami Y., Mochizuki M., Song C.,

RA Lee Y., Kai C., Takahashi E., Mikami T.;

RT "The complete DNA sequence and transcription map of the unique long

RT genome region of Marek's disease virus type 2";

RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

RN (2)

RP SEQUENCE FROM N.A.

RC SPECIES=Gallid herpesvirus 3; STRAIN=HPRS24;

RA Izumiya Y., Jang H., Ono M., Mikami T.;

RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.

RN (3)

RP SEQUENCE FROM N.A.

RC SPECIES=Gallid herpesvirus 3; STRAIN=HPRS24;

RA Izumiya Y., Jang H., Ono M., Mikami T.;

RT "A Complete Genomic DNA Sequence of Marek's Disease Virus Type 2,

RT Strain HPRS24";

RL Curr. Top. Microbiol. Immunol. 0:0-0 (2000);

DR EMBL; AB024414; BAA82925.1; -

DR EMBL; AB049735; BAB16539.1; -

DR InterPro; IPR000635; Viral_DNA_bind.

DR Pfam; PF00747; Viral_DNA_dp; 1.

SEQUENCE 1190 AA; 130357 MW; 05589C36CA20BDE9 CRC64;

Query Match 44.4%; Score 2793.5; DB 12; Length 1190;

Best Local Similarity 45.2%; Pred. No. 4; Ser-221;

Matches 545; Conservative 236; Mismatches 396; Indels 29; Gaps 13;

Qy 6 KTVTVTPGPGYVYACRVEDLDEISFLAARSTDSGLALLPLMRNLTVKEKTSFLAVV 55

Db 5 KSVKITGGPTGYVYAAPTSSMPADLDSIFAAKSNDCSDAILPLVSLGTVLEADFMNVAUV 54

Qy 66 SGARTTGLAGAGITLKLTHSHFYPSVFVFGKHVLPSSAAPNLTRACNAARERFGFSRC 125

Db 65 AGTKTTGLSGSTTLKLVPHYHPCVVFHGGECIKPTKAPNLTRACDLARGFGYS-T 123

Qy 126 QGPPVDGAVETGAEICTRIGLEPENTILYLVVTALEKAVFNCVNFVFLHYGGDIYHINH 185

Db 124 YSPPVATSFETTGEOICDNLGMNPQETMLVLTVELFKEAVYLCNSLYHYGGTVGSING 183

Qy 186 GDVIRIPLFPVQLEPMPDNRLVPDPFNTHRSIGEGFVYPTFPYNTGLCHLHDCVIAPM 245

Db 184 VDVRRILYPLHLVFPDNVADPSTKRALGEGALMPKAFVNSLCLLHGYVISTA 243

Qy 246 AVALRVNVTAVAGAAHLAFDENHEGAVLPDITYTYFOSSSGTITARGARNVDNST 305

Db 244 AVGLRVNVDIAICAAHLSDENHEGTLFPADTAFAFNAFTPIESASKSOYKAGKEGMEL 303

Db	654	LAVPYASGICPVAFUSRTTIAVQDMALSOCSIMQOQVARNFRFQFOVLKRRV	713
Qy	715	VDFNGGFISTRSITVTLSBQPSAPNPTLQDAPAGRTFDGLARVSVEVIRDIRVKNR	774
Db	714	LELQAGFITSKTITVLEQOICVDPDSQSDSVSNWEGDLVKVTVEIPRELKVKK	773
Qy	775	VPFSGNCTN-LSEARARLVGLASAYQQRKRVDMHLGALGFLFKQFHGLLPPRGMPNNS	833
Db	774	VLFQGGIAGAASEATKSRLAGVVEAYQRPTKTHVLNGLPGLGFAKSYHTLLFPDVRQVNG	833
Qy	834	KSPNPQFWTLQORNOPADKLTHEEITTTAAVKRPTTEEVAAINFNLPPTCIGELAQFY	893
Db	834	ATPNALWFILLNQLPAGILSKEEDKSLFKFKTKSYADWYINISPTCPDGLAQFY	893
Qy	894	MANILKYCDHSQVYLNTLSIITGARRPRDPSVLHWIRKOVTSADTQAKALLEKT	953
Db	894	LANTILKYCHKHFFINTISALVAVSRRPDPAIVLFWIERPITKQDVAPAAQQLIASM	953
Qy	954	ENLPFWTTAFTSTHLVRAAMNORPMVVLGCSISKYHGAAGNRRVFCAGNWSGLANGKV	1013
Db	954	SDKDIWCAFTSSNLVGSIMTKPFFVICISISKYHGMAGSTKVFQSGNWNIMGGRNV	1013
Qy	1014	CPITFDTRRFIIACPRGGFICVPTGPSSGNRETTLSQVIRGIIIVSGGAMVQLAIYATV	1073
Db	1014	CSLMSFDRTHRYVMTCPRVGFVAEQPIFSSGIXKSTTLIDRVNVLSEESAAPHAAYMLA	1073
2y	1074	VRAVGAQAHMADDMLSLTDDEFLARDLEELHDOIIQILETPTWVEGALBAVKILDEKT	1133
Db	1074	LKMGVDRVQMELEDMWEIINDEYISLDELKQ-VVEAGGNADAAMTAKEMVYNA	1132
Qy	1134	TAGDEPTNLAFNFCSPESHOTTNSVLNISGSNISGTVPGLKRPDDDELFDLSGIP	1193
Db	1133	MSIPTDQPT---PDDFACDENLGHADGQTISETN-----LKP--NMNVFDLEPIP	1179
2y	1194	IK 1195	
Db	1180	EK 1181	

RESULT 9

ID	QDPO8	PRELIMINARY;	PRT;	1190	AA.
AC	QDPO8;				
DT	01-MAR-2001	(T-EMBLrel. 16, Created)			
DT	01-MAR-2001	(T-EMBLrel. 16, Last sequence update)			
DT	01-DEC-2001	(T-EMBLrel. 19, Last annotation update)			
DE	UL29	single stranded DNA binding protein.			
3N	HYT037.				
DS	Melesgrid herpesvirus 1 (herpesvirus of turkeys).				
CC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;				
CC	Alphaherpesvirinae; Marek's disease-like viruses.				
CC	NCBI_TaxID=37108;				
CC	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=FC126;				
RC	MEDLINE=20578232; PubMed=1134310;				
RA	Afonso C.L., Tulman E.R., Lu Z., Zaak L., Rock D.L., Kutish G.F.;				
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.				
RT	"The genome of turkey herpesvirus";				
RL	J. Virol. 75:971-978(2001).				
RL	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=FC126;				
RA	Afonso C.L., Tulman E.R., Lu Z., Zaak L., Rock D.L., Kutish G.K.;				
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF291866; AAC45767.1;				
DR	InterPro; IPR000635; Viral_DNA_bind.				
DR	Pfam; PF00747; viral_DNA_bp_1				
DR	SEQUENCE 1190 AA; 130054 NW; B450CAP5FFFD928 CRC64;				

Query Match 44.2%; Score 2781.5; DB 12; Length 1190;
 Best Local Similarity 44.5%; Pred. No. 4.4e-220;
 Matches 538; Conservative 245; Mismatches 391; Indels 35; Gaps 14;

Qy	1	MENTQKTVTFTGVLGYACRVEDDLERISFLAARSTDSLDLALLPLMNLVETKFTTS	60
Db	1	MDGVKSVKLVGGIGVYVNSKDAVPDELSLFAAKCIDSSESLLELAGLVESEFVR	60
Qy	61	SLAVVSGARTTLAGAGITLKTTHSHYPVSVFPHGGRVLPSSAAPNLTRACNAARERF	120
Db	61	NVAAVAGTKTKLVNGGTTIKLVETHYHNVFVYGGDCIKPCSKAPNLTKACEYARERF	120
Qy	121	GFSCQCPVPDGAVENTTGABICTRLGLEPENTILYLVTALFKEAVFMCNVFHYGLDI	180
Db	121	CYCYSAP---GSVETSGKQICEQGLDACKTMYLVVTELFKEAIVLNSFLYGGSDS	177
Qy	181	VHINHGDIRIPLPPVQFMPDVMRLVDPFNTHKSIGEGFYVTFYNTG-SCHLIHDC	240
Db	178	VTINDAEVRIRIPLPLHLVLPDFNRVSNBPSRPRALGDGAYMPEAFFNDALCRLLGY	237
Qy	241	VIAPMAVALRVNVTAVARGAHLADENEGAVLPDITTYTF---CSSSGGTTARGAR	298
Db	238	VLGTDVAGLVRNIDAVARGAHLCPDENHEGILLPADTITFTATTPTADTGKQKQKAGK	297
Qy	299	RNDVNSTKPSPGGFERRLASIMAADTALHAEVIFMTGIYEETFTDIKEWPMFIGMEGT	358
Db	298	RDGAE----TQGGYERRTASLMASDATLAIENVIASVVEEPIPDVKKWPFCNPVGY	352
Qy	359	LPRNALGSTRVAVGICAMVSPNSALYLTVEVDSGMEAKDGG-PGSPNRFYQFAG	417
Db	353	TDRVEALSAYNGRVAGLVGAMVFNSSNVIYMTVGVASSSDGKESGVPAFPSYRFFQIAA	412
Qy	418	PHLAANPQTDGRDGHVLSQS---TGSSNTEFSDVLYALICGFGAPLALRLLFYLERCDA	473
Db	413	PHLAANPLVDRDGKPISGDALPKLAPSTSEYSLDVLILAGFCPOLLAFLFYLERCDG	472
Qy	474	GAFTHGHG-DALKVYGTTFDSEIIPCSICEKHTRPVCAHTTVHRLRORMPFGQATQPIG	532
Db	473	GANACHEDLDTVLVSSAIDADMPCELCDKTSVYCAHTTVKRELYRLEPKFGVQMGAMG	532
Qy	533	VFGTWSQYSDCDPLGNVAPYLILRKFGDQTEAAKATMODTYRATLERIFIDLEQERLID	592
Db	533	LFGMATNVCVVALGSYAQFSTLKR--SEGEASRSVMQDTYLTVERVMKALEKEGLLA	590
Qy	593	KGAPCSSEGLSSVIVDHTFRILDLRLARIEQTTPMKVLVETRDYKIREGLSEATHS	652
Db	591	CDDPMNMAPADASIKQGPSIHAVSTWRNIIIEGASQLARNLTIREYNIIEGLCDANHT	650
Qy	653	MALTFDPYSGAFPCITNLFVYKTRHLAVQDLALSQCHCYVYGOOEGRNFRQFQVLR	712
Db	651	LSLTIEPSSGFCFVLSFLARRTIIAVIQMALUSQCSMLMHGQGVARNRFTQFSVLR	710
Qy	713	RFVDLFNGGFISTRSITVTLSBQPSAPNPTLQDAPAGRTFDGLARVSVEVIRDIRYK	772
Db	711	RVLELQVAGFITKDIATULEDQHVAVPDPSPRSQYDPTVINMEGLDTRVTIELRELKVK	770
Qy	773	NRVVFSGNCT-NLSEARARLVGLASAYQQRKRVDMHLGALGFLKQFHGLLPPRGMP	831
Db	771	NRVIFGGVIGTASDAKRSANMIYAYQRTKMNVLNGLPLGFAIKRYHSALFNVVMP	830
Qy	832	NSKSPNQFWTLQORNOPADKLTHEEITTTAAVKRPTTEEVAAINFNLPPTCIGELAQ	891
Db	831	NGAVPNAHWFTLQORNOPADKLTHEEITTTAAVKRPTTEEVAAINFNLPPTCIGELAQ	890
Qy	892	PYMANILKYCDHSQVYLNTLSIITGARRPRDPSVLHWIRKOVTSADTQAKALLE	951
Db	891	PELANTILKYCHKHFFINTISALVAASKKPRDPAIVLPWIDKTIITQGRDVSIAAQLIG	950
Qy	952	KTNELPLWTTAFTSTHLVRAAMNORPMVVLGCSISKYHGAAGNRRVFCAGNWSGLNGK	1011
Db	951	TIADRKDAWCATFSATNLVGSVMSTPLVIGVSIKRYHGMAGSTKVFQSGNWNIMGGR	1010
Qy	1012	NVCPLFTFDRTRFIIACPRGGFICVPTGPSSGNRETTLSQVIRGIIIVSGGAMVQLAIYA	1071
Db	1011	NVCSLNGFDRTHRYVMAKPRVGVAEHTGFSGLKEATLMDRARAILSEGGGAPHAHVYM	1070
Qy	1072	TVRAVGARQAQHWAFDDMLSLTDDEFLARDLEELHDOIIQILETPTWVEGALBAVKILDE	1131

191 IPLPVLQLEWPDVN-RLVDPDPFNTHRSIGEGFVYPTFPYNTGLCHLHDCVIAVAVAL 249
 202 IPIPVQWMPDIALRICRNPDSNSNIGECIYPTPLFNKALNRLHATGLTGQGSGL 261
 250 RVNVTAVARGAHLAFDENHGAVALPPIITVYFSSSSGTTTARGARNVDNSTKPS 309
 262 RTNDEAVARGAHLAFDGSFEGCVLTSTKTTEL-----AQNIQQTGPPK 307
 310 PSGGFERRLASIAADTAHAEIFNTGTYESTPTD-----IKWPMFIGNEGTLPLN 363
 308 QNVEVEREACSLAELATLSVSCAY---PFDNGGGLPYEWPFLFCDCKTQWERVQ 364
 364 ALGSYTARVAGVIGAMVGFNPGSALYLYTEYDSGMTAKDGGSG-----PSFNRFYFAGPH 419
 365 ALDRFSAELAGVAGIAFNNSPLYASEVVDGAPD-----PGERASSGLTRFYLTGLH 419
 420 LAANPOTDRGH-----VLSSOSTGSSNTEFSDVIALICGFGAPLLARLLFYLERCDAG 474
 420 LIGCPQVDVSGKRWEGTSCPSLTPONTFEYDPEHLAYACGFSPELLSRIIFYLEKCSQ 479
 475 AFTGGHGDALKYVTGTFTFSEIFCS-----LCEKHTRPVCAHTTVHRLORMRPFQAT 527
 480 LSTRPN---MNLISINSKCACTGAVPECRKCNDESNNHCIRYTMQRLTRFPFPRMR 536
 528 ROPIGVFTMNSQYSDCDPLGNVAYLYILRKPDQTEAAKTMQTYRATLERLFDLEQ 587
 537 RSPFIAGVAGDGYTDCDILGNFAYSQLKAGD-GEPAKAVMNDYRGIGWRVFOYLV 595
 588 ERLDRGAPCSSEGSSVIVDHTFRRILDLARIEOTTCPMKVAVETDCKIREGLS 647
 596 EGLINKDGTEDTHNSSL-----AEKTVFKIONFVGSECKFISALSGVRAYHYKSHLS 651
 648 EATHSMALTFDYGAFCPITNPLVKRTHLAVQDLALSQCHVYFQQVQGRNFRNQFQ 707
 652 STVTFG-SPEYSSAYCPVLSLLCAQTRSLFDLILSLQIHGTFDTRQPSKMFSSAM 710
 708 PVLRERFVLDNGGISTRSITVILSEGVAPNPTLQDAPAGETDGDILARVSEVIR 767
 711 PALRAFNGMLDKGLSGKYEPFSSIVPCVNAPDTPMRENTQCAIQAQYSEYLSRQGVILKX 770
 768 DIRKRVVVFSGNCTNLSEAAARLVLASAYQKQEKVDMLHGAFLGKQPHGLLPPR 827
 771 EFKVKNRVFVNGSSFSNS-----788
 828 GMPNKSQPNQFWTLQORNQMPADKLTHEEITIAAVKFTFEYAAINFILPPTCIG 887
 789 -----788
 888 ELAQYMANLILKYCDHSQYLINTLTSITGARRPRDPSSVLHWIRKDVTSAADETQAK 947
 789 -----788
 948 ALLEKTENLPBLWTTAFTSTHLVRAAMNQPMVVLGIGISIKYHGAAGNRRVFOAGNWSGL 1007
 789 -----CRVLSKFPVLMGIMISKYVQOQNSTTVFQSGNWSGL 826
 1008 ---NGKNVCPLFTFDRTRRIIACRPGFTCPVTPGSSGNRETTLSQDVRGIIVSGAM 1064
 827 IGNSGQSNVSNALNDPVRKFPALACKRVGTLSSGQVASCQFQENTIASQVRSILDAGGS- 885
 1065 VQAIATVAVRAVAGARQAHMAFDWLSLTDDEFLARDLEELHDOIIQTLPTWVEGALE 1124
 886 -PTSITAVLVLEGHMKDITAEITMAITDDKYLINILVELREDISGS-QNGSVATAQS 943
 1125 AVKIDKETT 1134
 944 MLSELEGKDT 953

RESULT 13
 40913
 040913
 040913;

PRELIMINARY; PRT; 1132 AA.

DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE ORF 06.
 OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 OX NCBI_TaxID=37296;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=97296220; PubMed=9151804;
 RA Neipel F., Albrecht J.C., Fleckenstein B.;
 RA "Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus
 human herpesvirus 8: determinants of its pathogenicity";
 RL J. Virol. 71:4187-4192(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Neipel F., Albrecht J.-C., Ensser A., Huang Y.-Q., Li J.J.,
 RA Friedmann-Kien A.E., Fleckenstein B.;
 RA "The genome of human herpesvirus 8 cloned from Kaposi's sarcoma";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U93872; AAB62603.1;
 DR InterPro; IPR002114; Hpr Serp site.
 DR InterPro; IPR001993; Mitoch carrier.
 DR InterPro; IPR000635; Viral DNA_bind.
 DR Pfam; PF00747; Viral DNA_bp.1.
 DR PROSITE; PS00215; MITOCH_CARRIER.1.
 DR PROSITE; PS00589; PTS_HPR_SER.1.
 SQ SEQUENCE 1132 AA; 125394 MW; CF89E4F451919CBF CRC64;

Query Match 10.9%; Score 683.5; DB 12; Length 1132;
 Best Local Similarity 23.5%; Pred. No. 9e-47;
 Matches 301; Conservative 196; Mismatches 500; Indels 285; Gaps 53;

QY 11 PTGPGYVACVREDLDEEISFLAARSTDSLLPLMRNLTKETTSLLAVVSGART 70
 DB 18 PTGPGCYLYAYLTHFPICEASLLGNGYPEAKVFSPLHLGLTVESDPLNVKAVH--- 73
 QY 71 TGLAGAGITLTKTTHSHFVSFVPHGKHVLPSSAAPNLTRACNAARERFGRSCQGPV 130
 DB 74 KIDATASVKLTSH---REALVFNTHLFPFGKGLKLEKRESRELFQSTFVEQOH 131
 QY 131 DGAVETTGABICTRIGLEPEPTILYLVVLTALFKEAVFNCVFLHYGGLD1-----VH 182
 DB 132 KGTL--WSPACQPLPCANE-IFMAVIVTEGKERL-----YGGKLVFVPSQTPVH 180
 QY 183 INHGDVIRIPLFPVOLFMDVNLVDPDPNTHRSIGEGFVYPTFPYNTGLCHLHDCVI 242
 DB 181 IGEHQAFKILPYDEDLFGPSRAQEL-----CRFYNDISRYLHDSIF 222
 QY 243 APMAVALRVNVTAVARGAHLAFDENHGAVALPPIITVYF-QSSSGTGTARGARN 301
 DB 223 TGIACALRVKDVSTVIQ-ASRQF--VHDQYKIKLVQAKDPQOCASRGTDG----- 271
 QY 302 VNSTKSPSGGFERLERLASINAAOTLHAHVFNMTI-YEETPTDIKE-----WPMFIG 354
 DB 272 -----STLMVID-SLVAELCMGVGLSFIEGPQDSCEVLNYDTWPIFEN 313
 QY 355 MBGTLPRLNALGSTRVAGVIGAWFSPNSALYTEVDSGMTAKDGGGSPFNFYQ 414
 DB 314 CETPPARLALVHBAEALHIGALFAANSVLYLTRV--AKLPQKQVGDANNYSFYL 371
 QY 415 FAGPHLAANPQTRDG-----HVLSSOSTGSSNTEFSDVIALICGFGAPLLARLLFYLE 469
 DB 372 QHGLGYLSEATVKENGASAFKGVFVSALDGGSS---YTLQHLAYASSFPHLLARMCYLQ 428
 QY 470 RCDAGFTGGHG-DALKYVTCFTFSEIPCSCEKHTRPVCAMHTVHRLORMPR-FQAT 527
 DB 429 FLEHHKNTNSQSYNVVDYV-GTAAPSQMCDLCCQCCPACVNCINTLYFRMKDRFPPLSVNK 487
 QY 528 ROPIGVFTMNSQYSDCDPLGNVAYLYILRKPDQTEAAKATMDQTYRATLERLFDLEQ 587
 DB 488 RDPYVITGTACT-YNDLEILGNLPAFREREEGNVEDAP---KYTTWQLCQNI-----T 538


```

Db      859 ---PVPV-----QDLTYNDWTRDQSATVIQSSCRED-----PAAVCXMRPIVT 899
QY      982 LGISISKYHGAAGNNRVFOAGNWSGLNGKNVCPLFTFDR-----TRR 1024
Db      900 IPMMVNKYSNGNNSIFQSGN-MGYFMGRGV-----DRKLLDPIMRKXHVNTSMRRR 952
QY      1025 FIACPRGGFICPVTGPSSGNETTLSDOVR---GIIVSGGAMVQLAIYATVVRANGAR 1080
Db      953 YAFASPMTESL--LTQPGKGTVRTYQLEKTKIHHDIASGNABPE-NLVLELVKCLGHE 1009
QY      1081 AQHMAFDDWLSLTDDEFLARDLEELHQIQTLETFWTVEGAL--EAYKILDEKTTAGDG 1138
Db      1010 CATLTS-----DVEFYLGEPFMMVNDWVEVLETKASGCAFSSESQALLPTSSPGTS 1063
QY      1139 ETPTNLAFNFDSCPESHDTT 1158
Db      1064 ESFVEVTNIIITSLFEISVT 1083

```

Search completed: January 30, 2004, 13:14:03
 Job time : 52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

run on: January 30, 2004, 13:07:07 : Search time 17 Seconds
(without alignments)
3327.831 Million cell updates/sec

Title: US-09-769-699-2

Perfect score: 6294

Sequence: 1 MENTQKTVTPGPGYVYA.....DELPLSGIPKIGNITWEM 1203

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	6294	100.0	1204	1 DNEI_VZVD	P09246 varicella-z
2	3620.5	57.5	1209	1 DNEI_HSVB2	P28932 equine herp
3	3054	48.5	1196	1 DNEI_HSVIK	P17470 herpes simp
4	3049	48.4	1196	1 DNEI_HSVIF	P04296 herpes simp
5	3047	48.4	1196	1 DNEI_HSVIF	P04296 herpes simp
6	3015.5	47.9	1196	1 DNEI_HSVH2	P89452 herpes simp
7	3013	47.9	1197	1 DNEI_HSV2	P36384 herpes simp
8	2933	46.6	1186	1 DNEI_HSVB2	P12639 bovine herp
9	1094.5	17.4	375	1 DNEI_HSV21	Q03444 equine herp
10	669	10.6	1128	1 DNEI_EBV	P03227 Epstein-Bar
11	604	9.6	1128	1 DNEI_HSVB2	P24910 herpesvirus
12	492.5	7.8	1132	1 DNEI_HSV6U	P52338 human herpe
13	490.5	7.8	1132	1 DNEI_HSV6Z	P52538 human herpe
14	483	7.7	1131	1 DNEI_HSV7J	P52339 human herpe
15	479.5	7.6	1160	1 DNEI_SCMVC	P13215 simian cyto
16	477	7.6	1235	1 DNEI_HCMVC	P17147 human cytom
17	473	7.5	1193	1 DNEI_HSV72	Q04717 herpesvirus
18	412	6.5	1191	1 DNEI_HCMV2	Q04717 herpesvirus
19	121	1.9	1066	1 Z295_HUMAN	P30672 murine cyto
20	115	1.8	823	1 CRK3_CHICK	Q04173 homo sapien
21	114	1.8	822	1 IF2C_EUGER	P18461 gallus gall
22	114	1.8	1941	1 YRM8_CAEL	Q04173 gallus gall
23	113.5	1.8	1886	1 GP21_EAT	Q09417 caenorhabdi
24	113	1.8	2505	1 FAS_RAT	P11654 rattus norv
25	110	1.7	3333	1 LNA3_MOUSE	P12785 rattus norv
26	109	1.7	2845	1 APC_MOUSE	Q61789 mus musculu
27	108	1.7	827	1 XANP_XANS2	Q61315 mus musculu
28	108	1.7	1902	1 P2P_LACPA	Q02470 xanthomonas
29	107	1.7	1279	1 BGSC_PSERL	Q02470 lactobacill
30	105.5	1.7	3298	1 PC16_HUMAN	P28937 pseudomonas
31	105.5	1.7	3567	1 ERY2_SACER	Q06J90 homo sapien
32	106	1.7	773	1 PAC2_PSSB3	Q03132 saccharopol
33	106	1.7	3149	1 TSGU_EBV	P15558 pseudomonas
					P03186 epstein-bar

ALIGNMENTS

RESULT 1

ID	DNEI_VZVD	STANDARD;	PRT;	1204 AA.
AC	P09246;			
DT	01-MAR-1989 (Rel. 10, Created)			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Major DNA-binding protein.			
GN	29			
OS	Varicella-zoster virus (strain Dumas) (VZV).			
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;			
OC	Alphaherpesvirinae; Varicelloviruses.			
OX	NCBI_TaxID=10338;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=86306657; PubMed=3018124;			
RA	Davison A.G., Scott J.E.			
RF	The complete DNA sequence of varicella-zoster virus;			
RL	J. Gen. Virol. 67:1759-1816(1986).			
CC	-I- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA REPLICATION.			
CC	-I- SUBCELLULAR LOCATION: Nuclear (Probable).			
CC	-I- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN FAMILY.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)

EMBL; X04370; CAA27912.1; -
PIR; C27214; DNB29
InterPro; IPR000635; Viral_DNA_bind.
Pfam; PF00747; Viral_DNA_bp; 1
DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
ZNFING 497 510 C2HC-TYPE
FT
SEQUENCE 1204 AA; 132139 MW; D2FEFE65DC0C674 CRC64;

Query Match 100.0%; Score 6294; DB 1; Length 1204;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MENTQKTVTPGPGYVYACRVEDLDLEISPLAARSTDSDLAALLPLMRNLTVKEFTFS	60
DB	1	MENTQKTVTPGPGYVYACRVEDLDLEISPLAARSTDSDLAALLPLMRNLTVKEFTFS	60
QY	61	SLAVVSGARTTGLAGAGITLKTTSHPFVPSVFPHGKHVLPSSAANLFRACNAARERF	120
DB	61	SLAVVSGARTTGLAGAGITLKTTSHPFVPSVFPHGKHVLPSSAANLFRACNAARERF	120
QY	121	GFSCRQGPVVDGAVETTGAEICTRLGLEPENTILYLVVTALFKFAVFCNVFLHYGLDI	180
DB	121	GFSCRQGPVVDGAVETTGAEICTRLGLEPENTILYLVVTALFKFAVFCNVFLHYGLDI	180

09/769,699 212/04
Search Notes

181 VHINHGVDVIRIPPEVLFMPDVNRLVDPDPNTHSRSTGEGVYVPTPYNTGLCHLIHDC 240
181 VHINHGVDVIRIPPEVLFMPDVNRLVDPDPNTHSRSTGEGVYVPTPYNTGLCHLIHDC 240
241 VIAPMAVALRVNTAVARGAAHAFDENHEGAVLPDITTYTFQSSSGTATTARGARN 300
241 VIAPMAVALRVNTAVARGAAHAFDENHEGAVLPDITTYTFQSSSGTATTARGARN 300
301 DVNSTKSPSGGPERLASMAADTALHAENVFNTGIYEETPTDIKEWPMFIMGSTLP 360
301 DVNSTKSPSGGPERLASMAADTALHAENVFNTGIYEETPTDIKEWPMFIMGSTLP 360
361 RLNALGSYTRAVAGVIGAMVFPNSALYLTEVEDSGMTEAKDGGPSPNRFYQFAGPHL 420
361 RLNALGSYTRAVAGVIGAMVFPNSALYLTEVEDSGMTEAKDGGPSPNRFYQFAGPHL 420
421 AANPOTDRDGHVLSQSTGSNTSEFSDYDIALICGFGAPILLAKLLFVLRCDAGATGGH 480
421 AANPOTDRDGHVLSQSTGSNTSEFSDYDIALICGFGAPILLAKLLFVLRCDAGATGGH 480
481 GDALKYVTGTFDSEIPCSLCEKTRPVCAHTVHRLRQMRPFGQATROPIGVFGTMNSQ 540
481 GDALKYVTGTFDSEIPCSLCEKTRPVCAHTVHRLRQMRPFGQATROPIGVFGTMNSQ 540
541 YSCDPLGNVAPYLIIRKPGDQTEAAKATQDTRATLERFIDLQERLLDRGAPCSSE 600
541 YSCDPLGNVAPYLIIRKPGDQTEAAKATQDTRATLERFIDLQERLLDRGAPCSSE 600
601 GLSVIVDHTFRILLTLARIEQTTQPMKVLVETRYKIREGLSEATHSWALTFFDY 660
601 GLSVIVDHTFRILLTLARIEQTTQPMKVLVETRYKIREGLSEATHSWALTFFDY 660
661 SGAFCPITNPLVRLTHLAVVQDLALSCCHVCFYQQVEGRNFRNQPVLRFRFVDFLNG 720
661 SGAFCPITNPLVRLTHLAVVQDLALSCCHVCFYQQVEGRNFRNQPVLRFRFVDFLNG 720
721 GFISTASITVTLSEGPVSNPTLGDAPAGRTFDGLARVSVEVIRDRVKNRVFNSN 780
721 GFISTASITVTLSEGPVSNPTLGDAPAGRTFDGLARVSVEVIRDRVKNRVFNSN 780
781 CTNLSEAAARLVGLASAVORQEKRVDMHLGALFLKQFHGLLPPRGMPNPKSPNPOW 840
781 CTNLSEAAARLVGLASAVORQEKRVDMHLGALFLKQFHGLLPPRGMPNPKSPNPOW 840
841 FWTLLORNQMPADKLTHEEITITIAAVKRTTEVAAINFNLPTTCIGELAQFYMANLILK 900
841 FWTLLORNQMPADKLTHEEITITIAAVKRTTEVAAINFNLPTTCIGELAQFYMANLILK 900
901 YCDHSOYLINTLTSITGARRPRDPSVLHWRKOVTSAADETQAKALLEKTEVLPWL 960
901 YCDHSOYLINTLTSITGARRPRDPSVLHWRKOVTSAADETQAKALLEKTEVLPWL 960
961 TTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGKVCPLFTFD 1020
961 TTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNNRVFQAGNWSGLNGKVCPLFTFD 1020
1021 RTRERFIACPRGGFCVPTGSPSGNRETTLSQVIRGIIIVSGGAMVQLAIYATVRAVAG 1080
1021 RTRERFIACPRGGFCVPTGSPSGNRETTLSQVIRGIIIVSGGAMVQLAIYATVRAVAG 1080
1081 AQHMAFDDMLSTDDEFLARDLEELHDQIIQTLETPTWTEGALEAVKILDEKTTAGDGET 1140
1081 AQHMAFDDMLSTDDEFLARDLEELHDQIIQTLETPTWTEGALEAVKILDEKTTAGDGET 1140
1141 PTLNLFNPSCEBPSHTTGNVLNIGSNTSGSTVPGELKRPBDEBDFDLSGTPIKHGNIT 1200
1141 PTLNLFNPSCEBPSHTTGNVLNIGSNTSGSTVPGELKRPBDEBDFDLSGTPIKHGNIT 1200
1201 MEM 1203
1201 MEM 1203

RESULT 2
DNBI HSVEB STANDARD; PRT, 1209 AA.
AC P28932;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major DNA-binding protein.
GN 31.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OC NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1";
RL Virology 189:304-316 (1992).
CC -1- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: M86664; AAB02466.1;
CC PIR: E36798; DNEC4;
CC InterPro: IPR000635; Viral_DNA_bind.
CC Pfam: PF00747; Viral_DNA_bp_1.
CC DNA-binding; DNA replication, Zinc-finger, Nuclear protein.
CC ZN_FING 503 516 C2HC-TYPE.
CC SEQUENCE 1209 AA; 129982 MW; 1A728FB0448APE95 CRC64;
Query Match 57.5%; Score 3620.5; DB 1; Length 1209;
Best Local Similarity 57.1%; Pred. No. 47; 270;
Matches 698; Conservative 191; Mismatches 299; Indels 35; Gaps 14;
QY 1 MENTQKTVVPTPLGVVYACRVEDLDLEEISPLAARSTDSPLALPLMRNLTVKRTFS 60
DB 1 MESAPKTVSLPVSPLGVVYARQASLQTGTVSLTAARSVDLAVLPVIRGLTVQTFIT 60
QY 61 SLAVSGARTTGLAGAGITLKLTTSHFYPSPVFFHGGKHLFSSAAPNLTRACNAARERF 120
DB 61 NVAVVAGSKTGLGGTGTILKLTSPHFNFNAFVYGGSVIGASSNAPNLTRACNAARERF 120
QY 121 GPRSCGPPVDGAVETTGAEICTRLGLSPENILVLYVTALKEAVFNCVPLHVGGLDI 180
DB 121 GFSAFSSPPVENAVETSGEIEICASINLSPETLALYLVVTSFKEMVYVNTFLHYGGTST 180
QY 181 VHINHGVDVIRIPPEVLFMPDVNRLVDPDPNTHSRSTGEGVYVPTPYNTGLCHLIHDC 240
DB 181 VTIDGODAMKIPYVPLYWPNVRLASEPFFNAKHSIGDEYVSRPFNSDLCHLHGY 240
QY 241 VIAPMAVALRVNTAVARGAAHAFDENHEGAVLPDITTYTFQSSSGTATTARGARN 298
DB 241 VLGPAAVALRVNLDGARGAAHAFDENHEGAVLPDITTYTFQSSSGTATTARGARN 300
QY 299 RNDVNSTKSPSGGPERLASMAADTALHAENVFNTGIYEETPTDIKEWPMFIMGST 358
DB 301 QGD-GSGKNSASSGIERELASVMAADTALSVDSTWAGIYDTHELPSVEDWELVSSGDDT 359
QY 359 LPRINALGSYTRAVAGVIGAMVFPNSALYLTEVEDSGMTEAKDGGPSPNRFYQFAGPH 418
DB 360 -ESLEALGAYAAARLSGLVGAMVFSANVLYMTVEDDGGPADGKGS-NPSVHRFYLLAIP 417

```

419 HLAANPQTDGRDGHVL-----SSQSTGSSNTSEFSDYLYALICGFCAPILLARLLFYLBERCD 472
418 YVAGNFPQTDKORVLPHTADQQAAPINGSNQFSLDYALACGFCQFQALRLLFYLBERCD 477
473 AGAFTG-GHGDAKLYVTGTFDSEIPCSLCEKHTREVCARTVHLRQRMFRFGQATROPI 531
478 AGTEGGRNETDALRYLANLTESDVEGCLCNQATRPACATHTLHLRQLRFPFAPVRAP 537
532 GVEFTMNSQSDCLGNYPALILRKGDQTEAAKATMDTYATILBELFIDLEQRLL 591
538 GLEFTMNSAYSDDVLGNVASYGALKRND-NEAPKSIQMDTYATMRLELVNELEQAKLI 596
592 DR-----GAPCSSEGLSSVVDHPFRRLDTLRARIEQTTTQFMKVLVETRDYKIREGL 646
597 DKETLAQASPCGAP--TSVHVDQASFGILLNKNKTIEGAAEQFWTLVEARDKIREGL 654
647 SEATHSMALTEDPYSAGCPITNLFUKTHSLAVQDIALSQCCHVFGQVQGVGGRNFRNQF 706
655 ADANHTNSISLDPYSSSFCPTVSTFLARTVFAVLQDLVLSQCCHLFCYQSVGGRNFRNQF 714
707 QPVLRRRFRVDLFNGGFISTRITVTLSGPPVSAKNPTLQDAPAGRTFDGDLARVSVEVI 766
715 QPVLRRRFLDMLNGGFIKATVTVTVSDGVLADLTLRASEPTTKDYDGDGMARVSVEVL 774
767 RDIRVKNRVWFSNGMCTNLSEARARLVGLASAYORQEKRVQKRVDMHGLGFLIKQPHGLLFP 826
775 RDLRVKNRVLFNSGNGANMSEARARVAGMAGAYRRPKGNSILNGLVAGFLVKQVHGLVFP 834
827 RGMPPNSKSPQWFTLLQONQPADKLTBEITTTAAVKRFTTEVYAAINFNLPTPCI 886
835 RGHPPGIDTPNPQWFTLLQONQPADKLTBEITTTAAVKRFTTEVYAAINFNLPTPCI 894
897 GELAQFYMANILKYCDHSOYLNTLSITGARRPRDPSGLVHLWIRKDVTSADTETQA 946
895 GELAQFYFANLVKXCDHSQYFINGLTAVVGSRRPRDPAVLAWIDRTINGRADVEPAA 954
947 KALLEKTENLDELWTTAPTSTHLVRAAMQRPVVLGKISIKYHGAAGNNRVFQAGNWSG 1006
955 QEVLRQLGNSPAWMTGFTSTNMVRYMDQRPVVLGKISIKYHGAAGNNRVFQAGNWSG 1014
1007 LNGKGVCPLEFTRRRFIACPRGGFICPVTCPPSGNRETTLSQVRCIIYVSGGMVQ 1066
1015 LNGKGVCPLEFTRRRFIACPRGGFICPVTCPPSGNRETTLSQVRCIIYVSGGMVQ 1074
1067 LAIVATVVRVAGRAQHMADFDMLSLTDDFLARDLELHDOIIQITLTPWTVEGALEAV 1126
1075 TAVPAAVLHALGARTQHLAVDDWGLVDDEFLAASLDALNATVVDQF-GEWSVEAAQELV 1133
1127 KILDEKTTAG---DGETPTNLAFNFDSC---BPSHDTTSNVLNMGISGSGTVGLKQRP 1180
1134 KNEAQTTAGAVAAAGEG---AFDFGACVGTTPQOSTSA-----FNGGLAMAAAPAGQKRS 1185
1181 PEDDELFDLSGIPKIKHGNITMEM 1203
1186 LPDILDFMGAPPEKKSGLTFDM 1208

SUBUNIT 3
DDBI HSVIK STANDARD; PRT; 1196 AA.
P17470;
01-AUG-1990 (Rel. 15, Created)
01-AUG-1990 (Rel. 15, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Major DNA-binding protein (infected cell protein 8) (ICP 8 protein).
DAP OR UL29 OR ICP8.
Herpes simplex virus (type 1 / strain KOS).
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Alphaherpesvirinae; Simplexvirus.
NCBI_TaxID=10306;
[1]
SEQUENCE FROM N.A.

```

```

RX MEDLINE=88179536; PubMed=2833010;
RA Gao M., Bouchey J., Curtin K., Knipe D.M.;
RT "Genetic identification of a portion of the herpes simplex virus ICP8
RL Virolgy 163:319-329(1988).
CC -I- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION. LOCATION: NUCLEAR. IN THE ABSENCE OF DNA REPLICATION,
CC FOUND IN THE NUCLEAR FRAMEWORK-ASSOCIATED STRUCTURES
CC (PREREPLICATIVE SITES); AS VIRAL DNA REPLICATION PROCEEDS, IT
CC MIGRATES TO GLOBULAR INTRANUCLEAR STRUCTURES (REPLICATION
CC COMPARTMENTS).
CC -I- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M20165; AAA45793.1; .
CC PIR; A28601; DNBKKS.
CC InterPro: IPR000635; Viral DNA_bind.
CC Pfam: PF00747; Viral DNA_bp_1.
CC DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
CC ZN_FING 499 512 C2HC-TYPE.
CC SEQUENCE 1196 AA; 128314 MW; 0D0010A5DF0A4694 CRC64;
CC -----
Query Match 48.5%; Score 1054; DB 1; Length 1196;
Best Local Similarity 49.8%; Pred. No. 1.8e-256;
Matches 609; Conservative 200; Mismatches 366; Indels 48; Gaps 17;
CC 1 MENTOKTVT---VPTPLGYVYV--ACRVEDLDLBEISFLAARSTDSIALPLNRLTVE 55
CC 1 METKPKATTIKVPPGLPYVYVYACPPSEGIEL--LALLSARSGDADVAVAPLVGLTVE 58
CC 56 KTFITSSLAVSGARTTLAGAGITLKLTTSHFYSPVVFHGGKHLVSSAAPNLTACNA 115
CC 59 SGFEANVAVVGSRTTGLGGTAVSLKLTTPSHYSSSVYVFEGGRHLDESTQAPNLTLCER 118
CC 116 ARERFGRCQPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHY 175
CC 119 ARRHFGFSDYTPRGDLKHETTGEALCERGLDPRALLYLVTGFKAVCINNTFELH 178
CC 176 GGLDIVIHNGDVRIRLFPVQLFMPDVNRLVPPNTHRSIGEGFYVPTFPYNTGLCH 235
CC 179 GGSQKVTITGGAEVHRIPVYPLQLFMPDFSRVIAFPFNAHRSIGENFTYPLPFNRPLNR 238
CC 236 LIHDCVIAPMAVALRVNVTAVARGAAHLAFDENHGAFLPPDITVYFQSSSGTITAR 295
CC 239 LLFPAVVGPAVALRCNVDAVARAAHLAFDENHGAALPADITFTAFASQG--KTPR 296
CC 296 GARRNDVNSTKPSFGGFERRLASIMAADTALHAEVIFNTGIYEETPTDIKEWPMFIGM 355
CC 297 GGR-----DGGKGPAGGFEORLASVAGDAALAESIVSMVAFDEPPTDISAPLCEGQ 351
CC 356 EGTLPRLNALGYSYARVAGVIGAVMSPNSALYLTVEDESCMTAKDGGPGSPFNRYQF 415
CC 352 DTAARANAVAGYLARAAGLVGAVMSTNSALHLYTEVDDAGPADPKDHSK-PSFYRFFLV 410
CC 416 AGPHLAANPQTDGRDGHVL-----SSQSTGSSNTSEFSDYLYALICGFCAPILLARLLFYL 468
CC 411 PGTHVAANPQVDREGHVVPFEGRPAPLVGGTQ-EPAGEHLAMLGFGPALLAKMLFYL 469
CC 469 ERCDAGFTGGHG-DALKYVVTGTFDSEIPCSLCEKHTREVCARTVHLRQRMFRFGQAT 527
CC 470 ERCDGGVIVGRQEMDVPYVADSNQTDVPCNLCTFDTTRHACVHTTILMLRLASHPKFASAA 529
CC 528 RQPTGVFGTMNSOYSDCDPLGNVAPVILIRKPKDQTEAAKATMDTYATILBELFIDLEQ 587

```

Db 530 RGAIGVGTGMSMYSDCDVLGNVAFSAKLR-ADGSETARTIMQETVRAATERVMAELET 588
 Qy 588 ELLIDRGAPCSSEGLSSVVDHPTRRIIDTLRLARIEQTTQFMKVLVTRDYKIREGLS 647
 Db 589 LOYVDQAVFTAMGLETTIITREALTHTVNNVROVDREVEQLMRNLVEGRNFKFRDGLG 648
 Qy 648 EATHSMALTFDPYSGAFPIINFLVKRTHLAVQDLALSOCHCVFYGOQVEGNERNOFO 707
 Db 649 EANEANSITLDPYAGCPCLLOLGRSNLAVQDLALSOCHGVAFQSGVEGRNFRNOFO 708
 Qy 708 PVLRRFVLDLNGGFISTRSTVTLSEB-PVSAENPTLGDAPAGRTFDGDLARVSVEVI 766
 Db 709 PVLRRVMDMENGLSAKTLTVALSEGAICAPSLTAGTAPAESFEQDVARVTLGPP 768
 Qy 767 RDIVKRVVFSGNCTNLSEABARLVGLASAVORQKRVDMLHGLFLLKQFGLLPP 826
 Db 769 KELRVKSEVIFAGASANASEAKARVASQAKPKRVDIILLGLFLLKQFHAAIIP 828
 Qy 827 RGMPPSKSPNPQFWTLQORNPADKLTHEEITIAAKRFTVEYAAINFNLPTCI 886
 Db 829 NGKPPGSGNPQFWTALQORNPALSLGREDIETIAPIKFSLDYGAINFNLAPNV 888
 Qy 887 GELAQFYMANLILKYCDHSOYLNTLTSITGARRPRDPSVLHWIRKOVTSADLETQA 946
 Db 889 SELAMVYMAQILIRYCDHSYFINTLTITAGSRPPSVQAAAAW---SAQGGAGLEAGA 945
 Qy 947 KALLETENLPELWTAFTTHLVRAAMNORPMVLGISISKYHGAAGNNRVQAGNWSS 1006
 Db 946 RALMDAVIDHAGWTSFASCNLLRPVMAARPMVLGISISKYGVAGNDRVQAGNWAS 1005
 Qy 1007 LMGKQKVCPLFTDRTRFIIACPRGFCVPTGSPSGNRETLSDQVRGIIVSGAMVQ 1066
 Db 1006 LMGKQKACPLIFDRTRKFVLACPRAGFVCAASNLGGGAHESLCEQLRGIISEGGAVA 1065
 Qy 1067 LATVATVRAVAGARAOHMAVDDMLSLTDDDEFLLDLELHDLQIQLTLEPWTVEGAL--- 1123
 Db 1066 SSVFVATVKSGLRPTQQLQEDMLLEDLSEEMWELTARALRNGENWSTDALEVA 1125
 Qy 1124 -EAVKILDEKTAGDGETPTNLAFNFD--SCEPSHDTTGNVLNISGNSISGTVPLKAP 1180
 Db 1126 HEAEALVSOQGNAGE-----VFNFGDFCE-----DDNATPFGGFGAPGAPAGRKEA 1173
 Qy 1181 PEDELFDLSIGIPIKGNITWEM 1203
 Db 1174 FHGDPPFG-EGPPDKKGDLLDM 1195

RESULT 4

DNBI_HSV11 STANDARD; PRT; 1196 AA.
 AC P04236
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 20-MAR-1987 (Rel. 04, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Major DNA-binding protein (infected cell protein 8) (ICP 8 protein).
 DN DBP OR UL29 OR ICP8
 DC Herpes simplex virus (type 1 / strain 17).
 DC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 CC Alphaherpesvirinae; Simplexvirus.
 UN NCBI_TaxID=10292;
 UN [1]
 UN SEQUENCE FROM N.A.
 RP MEDLINE=88274327; PubMed=283594;
 RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
 RA McNab D., Perry L.J., Scott J.E., Taylor P.;
 RA "The complete DNA sequence of the long unique region in the genome of
 RT herpes simplex virus type 1";
 RT J. Gen. Virol. 69:1531-1574(1988).
 UN [2]
 UN SEQUENCE FROM N.A.
 RP MEDLINE=86067223; PubMed=2999714;
 RA Quinn J.F., McGeoch D.J.;
 RA "DNA sequence of the region in the genome of herpes simplex virus

RT type 1 containing the genes for DNA polymerase and the major DNA
 binding protein.";
 RL Nucleic Acids Res. 13:8143-8163(1985).
 RN [3]

RP SEQUENCE OF 1062-1196 FROM N.A.
 RX MEDLINE=88306232; PubMed=2457278;
 RA Hamerschmidt W., Conraths F., Mankertz J., Buhk H.-J., Pauli G.,
 RA Ludwig H.;

RT "Common epitopes of glycoprotein B map within the major DNA-binding
 RT proteins of bovine herpesvirus type 2 (BHV-2) and herpes simplex
 RT virus type 1 (HSV-1).";

RL Virology 165:406-418(1988)
 CC 1- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
 CC REPLICATION.

CC 1- SUBCELLULAR LOCATION: NUCLEAR. IN THE ABSENCE OF DNA REPLICATION,
 CC FOUND IN THE NUCLEAR FRAMEWORK-ASSOCIATED STRUCTURES
 CC (PERREPLICATIVE SITES); AS VIRAL DNA REPLICATION PROCEEDS, IT
 CC MIGRATES TO GLOULAR INTRANUCLEAR STRUCTURES (REPLICATION
 CC COMPARTMENTS).

CC 1- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
 CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL; X14112; CAA32322.1; --
 CC EMBL; X03181; CAA26940.1; --
 CC EMBL; M21631; AAA45787.1; --
 CC PIR; A03790; DNEBVI.

CC Interpro; IPR000635; Viral DNA_bind.
 CC Pfam; PF00747; viral DNA_bp; 1.
 CC DNA-binding, DNA replication, Zinc-finger; Nuclear protein.

FT ZN FING 499 512 C2HC-TVPE
 SQ SEQUENCE 1196 AA; 128349 MW; 453799162E5B99B9 CRC64;

Query Match 48.4%; Score 3049; DB 1; Length 1196;
 Best Local Similarity 49.8%; Pred. No. 4; 4e-226;
 Matches 609; Conservative 199; Mismatches 367; Indels 48; Gaps 17;

Qy 1 MENTQKTVT--VPTGGLGVY--ACRVEDLDEBTSLAARSTDSDLALLPLMRNLTV 55
 Db 1 MEYKPTATTIKVPPGPGVGVYARACPSGIEL--JALLSARSGSDVAVAPLVGLTV 58

Qy 56 KTFSSLAIVSGARTTLAGAGITLKTTSHPYPSYVFPHGKHLVLPSSAAPNLTRACNA 115
 Db 59 SGFEANVAVVGSRTTGLGTAVSLKLTSPSHYSSVYVPHGGRHLDPSTQAPNLRLCER 118

Qy 116 ARRRFGPSRCQGPVGVAVTGAETCTRLGLEPENTILYLVVTLAKFVAVCNVFLHY 175
 Db 119 ARRHFGSDVTPRGOLKHTTGEALCEGLDPRALLYLVTGFEKAVCNNTFLH 178

Qy 176 GGLDIVHINRGDVIRIFLFPVQLFMDVNLVDPDPNTHRSIGRIGGVYPTFYNTGLCH 235
 Db 179 GSGDKVTIGGAEVHRIPVFLQFLFMDPSKVLAEFPNANRHSIGKFTYPLFFPRLNR 238

Qy 236 LIHDCVIAFVAVLRVNTVAVARGAAHLAFDENHGAIVPDDIYVTPQSSSGSTTAR 295
 Db 239 LLEFVAVGPAVALRCRNVDAVARAAHLAFDENHGAALPADIITFAFASQG--KTTPR 296

Qy 296 GARRNDVNSTKSPSGGPFERRLASITMAADTALHAETVNTGIVETPTDKEWPMELCM 355
 Db 297 GGR----DGGKGAAGGFEORLASVWAGDALLALESIVMAVFEPTDTSAMPLEFGQ 352

Qy 356 EGTLPRLNALSITARVAVIGAMVFPSPNSALYLTVEDSGMTEAKGGPGSPFNRFYQF 415
 Db 352 DTAARANAVGAYLARAAGLVGAMVFPNSALHLTEVDGADPADPKHKK-PSFVRFFLV 410

Qy 416 AGPHLAANPQTDGHLV-----SSQSTGSSNTFSVDYIALICGFGAPILALLFYL 468

```
db 411 PGTHVAANPQVDRGHVPGFEGRETAPLVGGTQ-EFAGEHLAMLCGFSALLAKMLFYL 469
> 469 EBCDAGFTGGHG-DALKVVTGTFDFSEPCSLCEKHTRPVCAHTTTHLRLQRMRFEGAT 527
db 470 EBCDGAIVGDEMDFRVVADSNQDTPCNCTDTHACVHTTLMELRAHHPKFSAA 529
> 528 RQPIGVGTWMSQYSDCDPLGNYPYLILRKPGDTEAAKATMQDTYRATLERLFDLEQ 587
db 530 RGAIGVGTWMSQYSDCDPLGNYPYLILRKPGDTEAAKATMQDTYRATLERLFDLEQ 588
> 588 ERLDRGAPCSSEGLSSVVDHPTFRRLDTRARIEQTTFQMKVLYETDYKIREGLS 647
db 589 LQYVDAQVPTAMGRLETTITNREALHTVNNVRQVVDREVEQLMRLNVEGRNFRDGLG 648
> 648 EATHSMALTDFPYSGAFCPITNELVETHLVAVQDIALSOCHCVFVGGQVGEGRNFRNQF 707
db 649 EATHSMALTDFPYSGAFCPITNELVETHLVAVQDIALSOCHCVFVGGQVGEGRNFRNQF 708
> 708 PVLRRFRVDFLNGGFISTRSITVTLSEG-FVSAPNPTLQDAPAGRTFDDGLARVSVEVI 766
db 709 PVLRRFRVDFLNGGFISTRSITVTLSEG-FVSAPNPTLQDAPAGRTFDDGLARVSVEVI 768
> 767 RDIVKRVWFSNGCNLSSEARARLVGLAGAYOROEKRVDMHGLGALLKQFHGLFP 826
db 769 KELRVKSVLPAGASANASBAKARVASLQAYQKPKRVDILLOPLGLKQPHAAIFP 828
> 827 RGMPPNSKPNQFWTLQONQPADKLTHEITIAAVKRPTTEYAAINFINLPPTCI 886
db 829 NGKPPGNSQPNQFWTLQONQPADKLTHEITIAAVKRPTTEYAAINFINLPPTCI 888
> 887 GELACFYMANILKYCDHSQYLINTLSIIGARRPRDPSSVLHWIRKDVTSADISTQA 946
db 889 SELAMYNWAILKYCDHSQYLINTLSIIGARRPRDPSSVLHWIRKDVTSADISTQA 945
> 947 KALLEKTENPELMTTFTSTHLVRAAMNQRPMVVLGISIKYHGAAGNKNRVFOAGNWSG 1006
db 946 RALMDAVIDHAPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYVAGNDRVQAGNWS 1005
> 1007 LGGKNVCPLTFTORTRRFIACRGFIQVTPGPPSGNRETTLSQVRGIIVSGGAMVQ 1066
db 1006 LMGKNACPLTFTORTRRFIACRGFIQVTPGPPSGNRETTLSQVRGIIVSGGAMVQ 1065
> 1067 LAIYATVVRVAGARQWAFDMLSLTDDGFPLARDLEELHDQIIQTLETPTWTEGAL--- 1123
db 1066 SSVFVATVKSIGPTQQLQIEDWLALLEDEVLSEMMELTARALERNGENWSTDALEVA 1125
> 1124 -EAVKIIDEKTTAGGTPTNLAFND--SCEPSHDTTSNVLNISGNSIGSTVPLKRP 1180
db 1126 HEABALVSQNLAGE-----VENFGDFGCE-----DDNATPFGGPGAPGAPAGKRA 1173
> 1181 PEDDELFDLSGIPKIGNITWEM 1203
db 1174 FHGDDPFG-EGPPDKKGLDLDLM 1195
```

RESULT 5

```
DBI HSVIF STANDARD; PRF; 1196 AA.
> P17459.
01-AUG-1990 (Rel. 15, Created)
01-AUG-1990 (Rel. 15, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Major DNA-binding protein (Infected cell protein 8) (ICP 8 protein).
DBP OR UL29 OR ICP8.
Herpes simplex virus (type 1 / strain F).
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Alphaherpesvirinae; Simplexvirus.
NCBI TaxID=10304;
[1]
SEQUENCE FROM N.A.
MEDLINE=88306231; PubMed=2841793;
Hammerschmidt W., Conraths F., Mankertz J., Pauli G., Ludwig H.,
```

```
RA "Conservation of a gene cluster including glycoprotein B in bovine
RT herpesvirus type 2 (BHV-2) and herpes simplex virus type 1 (HSV-1).";
RL Virology 165:388-405(1988).
CC -/- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -/- SUBCELLULAR LOCATION: NUCLEAR. IN THE ABSENCE OF DNA REPLICATION,
CC FOUND IN THE NUCLEAR FRAMEWORK-ASSOCIATED STRUCTURES
CC (PREREPLICATIVE SITES); AS VIRAL DNA REPLICATION PROCEEDS, IT
CC MIGRATES TO GLOBULAR INTRANUCLEAR STRUCTURES (REPLICATION
CC COMPARTMENTS).
CC -/- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC FAMILY.
DR PIR; D29242; DNEBHF.
DR InterPro; IPR000635; Viral DNA_bind.
DR Pfam; PF00747; viral DNA_bp_1.
KW DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
FT ZN_FING 499 512
SQ SEQUENCE 1196 AA; 128373 MW; BC872584DDB1C8E2 CRC64;
```

```
Query Match 48.4%; Score 3047; DB 1; Length 1196;
Best Local Similarity 49.8%; Pred No. 6.3e-226;
Matches 609; Conservative 199; Mismatches 367; Indels 48; Gaps 17;

QY 1 MENTQKVT---VPTGPLYVY--ACRVEDLDLEISFLAARSTDSDLALLPLMNLIVE 55
Db 1 METPKTATTIKVPPGLGVYVARACPSEGIEL--LALLSARSGSDVAVAPLVGLIVE 58
QY 56 KFTSSLVAVSGARTTGLAGITILKLTTSHPYSVFPVHGKVLVPSAAPNLTRACNA 115
Db 59 SGFEANVAVVSGRTTGLGTAVSLKLTSPHSYSSVYVHGRLDPSQAPNLTRCER 118
QY 116 ARERFGRCQPPVDGAVETTGABICTRLGLEPENTILYLVVTALPKEAVFMCVNLFHY 175
Db 119 ARRHGFSDYTPRPGDLKHETTGALCERGLDPRALLYLVTEGKEAVINTFHL 178
QY 176 GGLDIVHNGDVIRIPLFVLQFMPOVNRVLPDPFNTHRSIGEGFYVPPFYNTGLCH 235
Db 179 GGSCKVTIIGAEVHRIPYVPLQFMFPDSRVIASFFNANHRSIGEKFTYPLPFNRPLNR 238
QY 236 LTHDCVIAPMAVALRVNRVAVAGAAHLAFDENHGAFLPDDITYTYFQSSSGTTAR 295
Db 239 LLEAVVCPAAVALRCNVDAVAAAHLAFDENHGAALPADITFTAFASQG--KTPR 296
QY 296 GARRNDVNSTKSPSPSGFERRLASIMAADTALHAETVNTGIYEETPTDIKEWPMFIGM 355
Db 297 GER-----DGGKGGAAGFEQRLASVMAGDAALAESIVSMVAFDEPPTDISAMFEGQ 351
QY 356 ECTLPLNALGYSYARVAGVIGAMVFSNLSALYLTVEDSCWTEAKDGGPGSPNRYQF 415
Db 352 DTAARANAVGAYLARAAGLVGAMVFSNTSALHTEVDADGAPADPKDHSK-PSFYRFLV 410
QY 416 AGPHLAANPQTRDRGHVL-----SSQSTGSSNTSPSDYALICGFGAPLARLLFYL 468
Db 411 PGTHVAANPQVDRGHVPGFEGRETAPLVGGTQ-EFAGEHLAMLCGFSALLAKMLFYL 469
QY 469 EBCDAGFTGGHG-DALKVVTGTFDFSEIPCSLCEKHTRPVCAHTTTHLRLQRMRFEGAT 527
Db 470 EBCDGAIVGDEMDFRVVADSNQDTPCNCTDTHACVHTTLMELRAHHPKFSAA 529
QY 528 RQPIGVGTWMSQYSDCDPLGNYPYLILRKPGDTEAAKATMQDTYRATLERLFDLEQ 587
Db 530 RGAIGVGTWMSQYSDCDPLGNYPYLILRKPGDTEAAKATMQDTYRATLERVMAELET 588
QY 588 ERLDRGAPCSSEGLSSVVDHPTFRRLDTRARIEQTTFQMKVLYETDYKIREGLS 647
Db 589 LQYVDAQVPTAMGRLETTITNREALHTVNNVRQVVDREVEQLMRLNVEGRNFKRDGLG 648
QY 648 EATHSMALTDFPYSGAFCPITNELVETHLVAVQDIALSOCHCVFVGGQVGEGRNFRNQF 707
Db 649 EATHSMALTDFPYSGAFCPITNELVETHLVAVQDIALSOCHCVFVGGQVGEGRNFRNQF 708
QY 708 PVLRRFRVDFLNGGFISTRSITVTLSEG-FVSAPNPTLQDAPAGRTFDDGLARVSVEVI 766
```

```

Db 709 PVLRRVDMFNGFLSAKTLTVALSEGAACASLTAQTAPRESSFEGDVARVTIGFP 769
QY 767 RDIRVKNRVSGNCTNUSEAARLVGLASAYQOEKRVDMHLGALGFLKQPHGLFP 826
Db 769 KELRVKSVLRFAGASANAESAARVASLQSAAYKPKRVKVDILLGFLGLKQPHAAIFP 828
QY 827 RGMPPNSKPNQPFWTLLQONQPADKLTHEEITTAAVKRPTEEVAAINFINLPPTCI 886
Db 829 NGKPPSGNQPNQPFWTALQONQPARLLSREDIETIAFKKFLDYGAINFINLAPNNV 888
QY 887 GELAQFYMANILKYCDHSOYLINTLSIITGARRPRDPSSVLLHWIRKDVTSADDETQA 946
Db 889 SELAMYYMANQILKYCDHSOYFINTLTALIAGRSPSPVQAAAW---SAQGGAGLEAG 945
QY 947 KALLEKTENLELWTTAFTSLHVRAMNQPMVILGISIKHGAAGNRVFOAGNWSG 1006
Db 946 RALVDVDAHFGAMTSPFASCNLLRPVMAARPMVVLGLSISKYGMAGNDRVFOAGNWS 1005
QY 1007 LMGKNVCPLFTFTRRFFIACPRGGFICPVTGPGSGNRETTLSDOVRGIIVSGGAMVQ 1066
Db 1006 LMGKNACPLIFTRTRKFLACPRAGFVCAASSLGGGAHESSICEQLRGIISGGAAVA 1065
QY 1067 LAIYATVVRAGARQAQHWADWLSLTDDEFLARDEBELHQIITLETPTWVEGAL--- 1123
Db 1066 SSVFVATVKSIGRPQTQLOIEDWLTALLEDEYLSSEMMELTARALRGNGEWSDDALEVA 1125
QY 1124 -EAVKILDEKTTAGDGTPTNLAFFNF--SCEPSHDTTSNVNLISGNTSGSVPGIKRP 1180
Db 1126 HEAEALVSQNLGAGE-----VFNDFDQCE-----DDNATPFGGAPGAPAFAGKKA 1173
QY 1181 PEDEFLDLSGIPKIKHGNITWEM 1203
Db 1174 FHGDDPFG-EGPDPKKGDLTDM 1195

RESULT 6
DNBI_HSV2H STANDARD; PRT; 1196 AA.
ID DNBI_HSV2H
AC P89452;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major DNA-binding protein (Infected cell protein 8) (ICP 8 protein).
DE DFP OR UL29 OR ICP8.
DE Herpes simplex virus (type 2 / strain HG52).
DE Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
DE Alphaherpesvirinae; Simplexvirus.
DE NCBI_TaxID=10315;
DE [1]
DE SEQUENCE FROM N.A.
DE Dolan A.;
DE Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DE -/- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
DE REPLICATION.
DE -/- SUBCELLULAR LOCATION: Nuclear (probable).
DE -/- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
DE FAMILY.
DE -----
DE This SWISS-PROT entry is copyright. It is produced through a collaboration
DE between the Swiss Institute of Bioinformatics and the EMBL outstation-
DE the European Bioinformatics Institute. There are no restrictions on its
DE use by non-profit institutions as long as its content is in no way
DE modified and this statement is not removed. Usage by and for commercial
DE entities requires a license agreement (See http://www.isb-sib.ch/announce/
DE or send an email to license@isb-sib.ch)
DE -----
DE EMBL; Z86099; CAB06754.1;
DE Intex-Pro; IPR000635; Viral DNA_bind.
DE Pfam; PF00747; viral DNA_bp; 1
DE DNA-binding; DNA replication; zinc-finger; Nuclear protein.
DE ZN_FING 499 512 C2HC-TYPS.
DE DOMAIN 1168 1196 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).

```

```

SQ SEQUENCE 1196 AA; 128423 MW; A19CA84328DD7F5 CRC64;
Query Match 47.9%; Score 3015.5; DB 1; Length 1196;
Best Local Similarity 49.0%; Pred No 1; 7e-223;
Matches 595; Conservative 213; Mismatches 366; Indels 41; Gaps 16;
QY 4 TQKTVTVPTGPIGVYV--ACRVEDLDIEEISFLAARSTDSLALLPLMRNLTVKTI TSS 61
Db 7 TTTTVKVPFGMGVYVGRACPAEGLEL--LSLLGARGGDADVAVAPLIVGLTVESGFAN 64
QY 62 LAVVSGARTTGLAGAGITLKLTTSHFVPSVFPFGKHLVLPSSAPNLTACNAAERFG 121
Db 65 VAAVVGSRITGLGTFVAVSLKMPSHYSVSVFVFGGRHLAPSTQAPNLTLCERARPHG 124
QY 122 FSRGCGPVDGAVETTGAEICTRLGLEPENTILYLVVLTALFKEAVFMCNLFVHLYGGDIV 181
Db 125 PADYAPRPCDLKHETTGDCALCERLGLDPRALLYLVIETGFREAVCISNTFLHGGWDKV 184
QY 182 HINHGDIVIRIPLFPVOLFMPPVNRVLPDPNTHRSIGEOPVPTPTPYNTGLCHLIHDCV 241
Db 185 TIGDAEVHRIPIVPLQMPDPFSRVIAADPNCNHRISGENFNYPPLFPFNRLPARLLFEAV 244
QY 242 TAPMAVALRVNRVNTAVARGAAHLAFDENHGAIVLPDITVTYFQSSSSGTTTARGARRND 301
Db 245 VGRAVALRAENVDAVARAAHLAFDENHGAALPADITETAPASOG--KPORGAR---299
QY 302 VNSTKSPSPGCPERRIASIMADTALHAENVNTGIYEETPTDIKEWPMFIMGEGTLP 361
Db 300 --DAGNKGPAQGFQRLASVMAGDAALALESIYSMAVEDPDPDITTWPLLEGOETPAAR 357
QY 362 LNALGSVTARVAGVIGAMVSPNSALYLTEVDSGMEAKDGGGSPSNRFYQFAPGPHLA 421
Db 358 AGAVGAYLARAAGLVGAMVFTNSALHLTEVDGAPADPKDHSK--PSFYRFFLVPGTHVA 416
QY 422 ANPQTRDGHVL-----SSQSTGSSNTEFSVDYDLALICGFGAPLIARLLFYLERCDAG 474
Db 417 ANPQLDREGHVFPYEGRPTAPLVGGTQ--EFAGEHLAMLGFGSPALLAKMLFYLERCDGG 475
QY 475 AFTGGHG--DALKYVTGTFDSEIPECSICEKHTRPVCAHTTVHRLRQMRPFGQATPIGV 533
Db 476 VIVORQEMDVFRVAVSGQTDVPCNLCCTFETRHACATTLMLRARSHPKFPASAARIGV 535
QY 534 FGTWNSQYSCDPLGNYPYLIIRKPGDQTEAATAQWQDTYRATLERLIDLEQERLLDR 593
Db 536 FGTWNSAYSCDVLGNYPYALSKR--ADGSENRTTQMSTYRATYRVAEAEALQYVDQ 594
QY 594 GAPCSSGLSSVIVDPTFRILDLARIEQTTQFMKVLVETRDYKIREGLSEATHSM 653
Db 595 AVPTALGRLETTIIGNREALHTVNNIKQLVREVBQLEMLT:EGRNFKFRDGLAEANHAM 654
QY 654 ALTFDPYSGAFCEPITNFLVKTHTLAVVQDLALSOCHCVFYGOQVGRNFRNQFPVLRRR 713
Db 655 SLSLDPYTCGPCPLQLLARRSNLAVYQDLALSOCHGVFAGOSVEGRNFRNQFPVLRRR 714
QY 714 FVLDLFGGFTSTRITVTLSEGG--PVSAPNPTIGQDAPAGRTFGDLARVSVSEVIRVK 772
Db 715 VMOLFNGFLSAKTLTVALSEGAACIAPSLTAGTAPABSSFEGDVARVTLGFKELRVK 774
QY 773 NRVVFSNCNTINSEAAARLVGLASAYQOEKRVDMHLGALGFLKQPHGLFPFGMPEN 832
Db 775 SRVLFAQASNASEAARVASLQSAAYKPKRVKVDILLGFLGLKQPHAVTFPFGKPPG 834
QY 833 SKSPNQFWMTLQONQPADKLTHEEITTAARVFTTEVAAINFINLPPTCIGELAQF 892
Db 835 SNQPNQFQEWFTALORNOLPARLLSREDIETIAFKKFLDYGAINFINLAPNNVSELA 894
QY 893 YMANLILKYCDHSOYLINTLSIITGARRPRDPSSVLLHWIRKDVTSADDETQAALLEK 952
Db 895 YMANQILKYCDHSOYFINTLTALIAGRSPSPVQAAAWAQ--GGAGLSAGARALMDS 951
QY 953 TENLPBLWTTAFTSLHVRAMNQPMVILGISIKHGAAGNRVFOAGNWSGNGKN 1012
Db 952 LDAHPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYGMAGNDRVFOAGNWSLLGKN 1011

```

[illegible]

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major DNA-binding protein.
GN DBP OR UL29.
OS Bovine herpesvirus type 2 (strain BMV) (Bovine mammillitis virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10296;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88306231; PubMed=2841793;
RA Hamerschmidt W., Conraths F., Mankertz J., Pauli G., Ludwig H.,
RA Buhk H.-J.;
RT "Conservation of a gene cluster including glycoprotein B in bovine
RT herpesvirus type 2 (BHV-2) and herpes simplex virus type 1 (HSV-1).";
RL Virology 165:388-405(1988).
RN [2]
RP SEQUENCE OF 1058-1186 FROM N.A.
RX MEDLINE=88306232; PubMed=2457278;
RA Hamerschmidt W., Conraths F., Mankertz J., Buhk H.-J., Pauli G.,
RA Ludwig H.;
RT "Common epitopes of glycoprotein B map within the major DNA-binding
RT proteins of bovine herpesvirus type 2 (BHV-2) and herpes simplex
RT virus type 1 (HSV-1).";
RL Virology 165:406-418(1988).
CC -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M21630; AAA46051.1; --
CC F1R; A29442; DNBEGB.
CC InterPro; IPR000635; Viral DNA_bind.
CC Pfam; PF00747; Viral DNA_bp.1.
CC DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
CC ZN FING 495 508 C2HC-TYPE
CC SEQUENCE 1186 AA; 127286 MW; A586ECC1479FED2C CRC64;
CC -----
CC Query Match 46.6%; Score 2933; DB 1; Length 1186;
CC Best Local Similarity 48.7%; Pred. No. 3.7e-217;
CC Matches 585; Conservative 188; Mismatches 373; Indels 56; Gaps 11;
CC
CC 1 MENTOKTVT---VPTGPGVYACRVEDLDLEISFLAARSTDSDLALPLMRNLTVERT 57
CC 1 MENKQKATTVKSPGPGLVYARRLPPEGLTELALLSARSADSDTAVLPLIAGLTVESG 60
CC
CC 58 FTSSLAVSGARTTLAGAGITLKLATSHFPYVFVHGKXVLPSSAAPNLTRACNAAR 117
CC 61 FDVNVAVVSGRTTGGTGVSLKMPESHVAPGAYVFGHRLHAPSSAAPNLSILCDRAR 120
CC
CC 118 ERFGESGCOFPVDGAVETTCATLCTRLGLEPNTILYLVVTLFKEAVFMCNVPFLHYGG 177
CC 121 VQGFSSFKPCCEAGETTGELGCHLGLNPNESLLYVIAEGFKEAVYISNTILHMG 180
CC
CC 178 LDIVHNGDIVIRIPLFPVQLEMPDVRNLRVPDPFNTHRSIGEGFVYPPFVNTGLCHLI 237
CC 181 VGVTVIAGEVRIPIYPLQMFNPDYCRVADPFNDNRHRAIGEFAYPDPFFNAKLASLL 240
CC
CC 238 HCVIAPMAVALRVNVTAVARGAHLAFDENHEGAVLPDITYTFOSSSGTTTARGA 297
CC 241 FGAAGVPAVALRNVDAVARAAHLAFDENHEGAALPADITTFADPTGS-----KAG 295
CC
CC 298 RRNDVNSTKSPSGGFRRLASIMAADTALHAENVTFGIYEETPTDKIEWPMFIMEG 357
CC 296 HRNPRC-----GGGFQRLASVMAGDAALALEIMSNVAFEEPPTDGTWPLTQCES 349

QY 358 TLPRNLALGSYARVAGVIGAMVFGSPNSALYTEVEDSGMTAKDGGPGSPNRFYQFAG 417
DB 350 TAARAAISGAYILGRAAGVIGAMVFGSPNSALHTEVDDAGPADPKD-PTKPSFYFFELVPG 408
QY 418 PHLAANPQDRDGHVLSQS-----TGSSNTEFSDVIALICGFGAPLLARLLFYLERC 471
DB 409 TVVAANPQDRDGRVVGAGHEGRIPIVVGGNHEFTCEHLATLCGFSPELLAKMLYLERC 468
QY 472 DAGAFTG-GHGDALKYVTGTDTSEIPCSLCEKHDPVCAHTTVHLRORMPRFQATQRP 530
DB 469 DSGVILGREMDTKVSDSAHTDVPCCLSLDNHSCHHTLLRLRAHHPKFTSTTRGA 528
QY 531 IGVFGTMSQYSCDPLGNYPYLILRKPGDTEAAKATMDTYRATLERLFIIDLEERL 590
DB 529 IGI FGVMASAYSDCDVLGNYSFSAIKEM-DVQETARAIMQETYSRVERVMAELINLY 587
QY 591 LDRGAPCSSEGLSSVIVDHPTRFRITLRIEIQTTQFMKVLVETDYKIRGLSEAT 650
DB 588 IDAAVPTSPAKLESITOREALQTVVSNVQVQVGEVAQLMRALVEGRGFRFRALGEAN 647
QY 651 HSMALTDPYSGAFCPITNPLVYKTHLAVQDLALSQCHCVYQQVQVEGRNFRNQFVL 710
DB 648 HANSITLDPHASVPCGLQMLGRSNLAVYQDLALSQCHGVFEQVGEQVGRNFRSQFVL 707
QY 711 RRRFVLDLNGFISTRSTVTLSEGP-VSAPNPTLQDAPAGRTFDGDLARVSVEVIRDI 769
DB 708 RRRVLDLNNNGFLSARTLVALTDGACISAPGVSGGHAASGFEQVGRNFRNQFVKEL 767
QY 770 RYKRVVFSNGCTNLSEARARLVGLASAYOROKRVDMLHGLGALLKQFHGLLPRGM 829
DB 768 RVKSVLFPAGAGPAASEAARARIAGLOSAYOKSKRVDLILPLGLFMKQFHALFPNGK 827
QY 830 PPNKSNPQWFTLLQONQMPADKLTHEITTTAAVKRTTEYVAANFINLPTCTIGEL 889
DB 828 PPGSDNPQWFTALQONQMPADKLTHEITTTAAVKRTTEYVAANFINLPTCTIGEL 887
QY 890 AQFYMANILIKYCDHSOVLINTLTSITGARRPRDPSVLHWIRKQVTSADLETQAKAL 949
DB 888 AMYVMAQILKYCDHSYFINTLITAGSRFPNQAANAAPRGOT---ELEAQRSV 944
QY 950 LEKTENIPELMTTFTSTHLVRAAMNORPMVVLGIGISKYHGAAGNNRVFQAGNWSGLNG 1009
DB 945 VANFGDHPGATWTFASCNLLRPVMAIRPMVVLGIGISKYHGAAGNNRVFQAGNLANLLG 1004
QY 1010 GKNVCLPFTDRTEREIIACPGGFCPVTPGSGNRETTLSQVREGIIVSGAMTQALAI 1069
DB 1005 GKNACPLIITDRTRKFIACPRAGFVCAAVSAGSAGHESLCEQLRAIIAEGGATVASDV 1064
QY 1070 YATVVRVAGARQHMADFDDWLSLTDDEFLARDLEELHQIIOLETPWTVGEALEAVKIL 1129
DB 1065 FAAAASLIGARVOQLQIEDWLALLEDEYLSSENMELAGRALERGGGWSLDAALDVAREA 1124
QY 1130 DEKTT-----AGDGETTNLAFNDS-----CEPSSHOTTGN 1160
DB 1125 EMVTRHVDABETTFDGFABDGFADAGLAVHLQSRRLPLACSDLFGDAPAEKNDLTLD 1184
QY 1161 VL 1162
DB 1185 ML 1186

RESULT 9
DNBI HSV1
ID DNBI HSV1 STANDARD; PRT; 375 AA.
AC Q0344;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major DNA-binding protein (Fragment).
GN 31.
OS Equine herpesvirus type 1 (isolate HV525A) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;


```

QY 520 MPREGATROPVIGFTGWNQSOYSCDPLGNVAPYLII-----RKPGDTEAAKATMQDYR 575
DB 476 PPPVMSQRRDPYVVSAGSGYNSTDFLGNFLNFDKEDDQRPDPDEPRYTYWQLNQNL 535
QY 576 ATLRLFLIDLEQERLLDRGAFCSSEGLSVIVDHTFRRLIDTLRLARIQTTTQFMKVLV 635
DB 536 ERLSRLGIDAEKLEKPHGP-----RDFVMEKQVDAVDAEVQFVWNSMA 582
QY 636 ETRDYKIREGLSEATHSNALTFDPSGAFCPINFLVRLTHLWVQDLALSQCCHVYQG 695
DB 583 K-NNITYKDLVKSCYHVNQVSCNFPQACPIFTQLFYRLSLTILQDISLPIQWC----- 636
QY 696 QVEGNEFRNQFPVLRFRFVDLENGGF-----ISRSITV-----TLSEGPVSAPNLTG- 745
DB 637 -----YEND-NPGLGOSPENLKGHTCTLCTFRSLAIDKGVLTAKAVHGEPTCDL 689
QY 746 -----QDAPAGRTFGDRLAVSVSEVIRIRVKNRVVFGNCTNISEARARLVGLASA 798
DB 690 PDLDAALQGRVYGRRLVRVMSKVLMLCPRNKIKNRVWFTGENAALQNS-----FIKS 742
QY 799 YQROEKRVDMHGLGALLKQFGLLFPFCMPNPKSPNPQWFWLLQRMQPADK-LTH 857
DB 743 TTRRENVII--INGPYMKFLNTYKTLF-----PDTKLSSLYLWHNFSRRSRVPVPSGASA 795
QY 858 EBITTIAA-VKRTEEYAAINFLPPTCIGELAQFYMANLILKYCDHSQYLINTLTSII 916
DB 796 EYSDLALFVDGGSRAHEESNVIDVVGNLVTVAKQRLNAILKACGQTFYISLIQGLV 855
QY 917 --TGARRPRDPSVLHWRKDVTSRAIDETQAKALLEKTENIPELWTTAFTTHLVRAAM 974
DB 856 PRQSVAPADYPHVLG--TRAVESAA-----AYAEATSSLTAT-TVVCAATDCLSQVC 905
QY 975 NORPMVVLGISISKYHGAAGNNRVFOAGNWSGLNGKVCPLFTFDRTRRFTIACPRGG- 1033
DB 906 KARPVVTLPTVINKYTGNGNNQIFQAGN-LGYFMGRGV-----DRN---LLQAPGAGL 955
QY 1034 -----FICPTVGPSSGNR-ETTLSDQVGRGIIVSGGAV-----QLAIYATV 1073
DB 956 RKAQSGSSMRKKFVFATPTGLTVKRTQATYETIENIRAGLEAIIQKQBEDCVFVW 1015
QY 1074 VRVAGARAQHMADFDDWLSLT--DDEFLARDLEELHDIQIOTLET-----PWTVEGALEA 1125
DB 1016 CNLVDAMGACA-----SLTRDDAEVLLGRFVLDVSLVETLATTASSGIEWTAAADF 1070
QY 1126 VKILDEKTTAGGETTNIAFNFDSCPESHDTTNVLNITSGNISG-----STVFGL 1177
DB 1071 L-----EGVWGGGCAQDNFISVAEPVSTASQASAGLLGGGGGGGGRKRRLATVLPGL 1126
QY 1178 K 1178
DB 1127 B 1127

```

RESULT 11

```

DDBI_HSVSA
ID DDBI_HSVSA STANDARD; PRT; 1128 AA.
AC P24910.
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major DNA-binding protein.
EN 6 OR KRRFL.
OS Herpesvirus saimiri (strain 11).
SC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
XC Gamaherpesvirinae; Rhadinovirus.
NCBI_TaxID=10383;
[1]
SEQUENCE FROM N.A.
IP MEDLINE=90163221; PubMed=2154888;
IC Albrecht J.-C., Fleckenstein B.;
A "Structural organization of the conserved gene block of Herpesvirus
T saimiri coding for DNA polymerase, Glycoprotein B, and major DNA

```

```

RT binding protein.";
RL Virology 174:533-542 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9333688; PubMed=1321287;
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RT "Primary structure of the herpesvirus saimiri genome.";
RL J. Virol. 66:5047-5058 (1992).
RN [3]
RP SEQUENCE OF 1-81 FROM N.A.
RX MEDLINE=9120588; PubMed=1850023;
RA Nicholas J., Coles L.S., Newman C., Honess R.W.;
RT "Regulation of the herpesvirus saimiri (HVS) delayed-early 110-
RT kilodalton promoter by HVS immediate-early gene products and a
RT homolog of the Epstein-Barr virus R trans activator.";
RL J. Virol. 65:2457-2466 (1991).
CC 1- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC 2- SUBCELLULAR LOCATION: Nuclear (Probable).
CC 3- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X64346; CAA45629.1; --
CC EMBL; M31122; AAA46162.1; --
CC EMBL; M60849; AAA46157.1; --
CC InterPro; IPR000635; Viral DNA bind.
CC Pfam; PF00747; Viral DNA bp; 1.
CC DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
CC ZN_FING 453 466
CC SEQUENCE 1128 AA; 127457 MW; BAADE861381328D CRC64;

```

Query Match

9.6%; Score 604; DB 1; Length 1128;

Best Local Similarity 22.9%; Pred.No.5.9e-38;

Matches 275; Conservative 189; Mismatches 509; Indels 226; Gaps 40;

QY 11 PTGPLGVACRVEDLDLEISFLAARSTDSLALPLMRNLTVKFTTSSLAIVSGART 70

DB 18 PVPECGIIVYPKEGPPFKESALLGNKNVGSAMSLPLSLDTVESNFSFNKAVHKID 77

QY 71 TGLAGAGITLKLTTSHFYPSVVFVFGKHVLPSSAAPNLTRACNAARERFGRSQGPPV 130

DB 78 M-----TLLVRVSAYHREAIVFFNTDLFEIFVGPGLDILCSARSILFYTNF--VPR 129

QY 131 DGAVETTGAEICTRIGLEPENILVLTALPKFAVFCMN-VFLHYGGLDIVIHNGDVI 189

DB 130 TDLRDTVDIKOLYAPPYSEDSCFMAVWVTEGFKERLYFGNLVPIIAQGLK-VQINGREAV 188

QY 190 RIPLFPVQLFMPDVNRLVPDPENTHRSIGEGVYPTPTFNTGLCHLHDCVIAPMVAL 249

DB 189 KIPLXDEDLFSKSHL-----PRFYTPSVSKYLHDSVFTSIAQAL 229

QY 250 RVNVTAVARGAAHLAFDENHEGAVLPDITVYFQSSSGTTTARGARRNDVNSTKPS 309

DB 230 RINDVESVIRASEKSIQDQYKLAKI-----VNSK----- 259

QY 310 PSGGFERRLASIWAADTA-----LHAEIFNTGI-YEETP-----TDIKEWPMFIGM 355

DB 260 -----DFSLSQVKQDASAFNWDICIAELAMSICYLSFLEAPQDCAVLDTYSWIFETA 314

QY 356 EGTLPRLNALGYSYARVAGVIGAMVFSNLSALYTEVEDSGMTEAKGGPGPSFNRYQF 415

DB 315 ETEEDRIKAIDQWNAAMSVHYVTHLFTNSVLYLTINKQTSKNKSEQN---VYNTYFMQ 371

```
2Y 416 AGHLLAANPQTRDGHVLSQSSTNTSFFVDYLALICGFGAPLLARLLFYLERCDAGA 475
b 372 HGLAADAATQRENGEPAPSGAPKSGGGTYTLHLALASSFPHLLANCYMQFCQHK 431
2Y 476 FTGGHGDALKYVTGFDSEIPSLCEKHTRPVCAHTVHRLQRWPR-FGQATROPIGVF 534
b 432 STINANYSVQVGTAAASDLCELCQGTCPASCINTFLYRLKDRPPVGLSGQRDPYVVT 491
Y 535 GTMSQYSCDPLGNVAPYLIIRKPGDQTEAAKATMOTYRATLERLIDLEQERLLDRG 594
b 492 G-VSQYNDLMLGNFAF---REKEDAVQNAESEKTYWQIQNVV-----EKLSTMG 542
Y 595 APCSEGLSSVIVDPTFRRLIDTLARIEQTTQFMKVLVETRDYKIREGLSEATHSMA 654
b 543 VTEGTVG-SELITDIOSELKTPRDINDVVDSEVKFMCLVK-NININPRETIKTVHVLH 600
Y 655 LTFDYSYGAFCPTNFLVKRTHLAVVQDLA-----SQCHCVFYQQOV 697
b 601 YCCNVFWQAPCAMFNLFYKSVLAIQIDCIPIAMTYEQDNPS:GMPSEWLKVHY--QT 658
Y 698 EGRNPR-----NQOPVLRRRRVDLFGGFISTRGITVTLSEGPVSAAPNPLGQ 746
b 659 IWTNEKSSCLDRGLVTSEHKIVHTDMFCDELN-----IDSALS-----GQ 699
Y 747 DAPAGRTEDGLARVSVEVIRDIRVKNRVFSGNCTNLSEARARLVGLASAYORQEKRV 806
b 700 IVP--MKMQVRLAKALLTPKTIKKNRIVFSN--SSMTETIQSGFIKSAT-----KKDS 750
Y 807 DMLGALGFLKQFHLGPRGMPENSKSPNPFOWTLLQNMOPA-DKLTHEBITTIA- 864
b 751 YIVTGPYKFLNSLHKWF-----PNAKISALYIWHFTSQKKQPLVLPGISRENWELAN 805
Y 865 AVKPTBEYAAININPPTCIGELAPYMANILKYCDHSQYLINTLSI-----ITG 918
b 806 YVETSSKMDHMDVLDIPTLLTAYAKVRLNNTILRTCGGTQFVATTLOCLLPLQTISA 865
Y 919 ARRRP---DPS--SVLHWIR--KD-----VISAADIETQAKALEKTENBELMTTAF 964
b 866 TEYPHVLLDQSIMVDHLSIKDKHALTVQTTLKEDIATVGK----- 908
Y 965 TSTHLVRAAMQRMVVLGISISKYHGAAGNRRVFOAGN---WSGLNGKNVCPLFT-FD 1020
b 909 -----QREIVTVPLVWNYKTGNGNTQIQFCGNLGYFNGRVDENLIPDSTGR 957
Y 1021 RTRFLIACRGGFICP-----VTGPPSGHRET-----TLDQVRGIIVSGAVQIAIYA 1071
b 958 RQNNSSWYRRHVRHVTMPVAHVKKSNLNNLTPEVETIRKNVQNFEDKD---NLNIFD 1014
Y 1072 TV-----VRVAGARQAHMAFDWLSLTDDEFLARDLELHDIQITLETPTVEGALEAV 1126
b 1015 NVVLELVKGLGDSCEINTEDDLOFYLGYYVIMSDEIWSRQIITDSGAPMSVENVTKVL 1073
3SUULT 12
b1 HSV6U STANDARD; PRT; 1132 AA.
b P52338;
b 01-OCT-1996 (Rel. 34, Created)
b 01-OCT-1996 (Rel. 34, Last sequence update)
b 16-OCT-2001 (Rel. 40, Last annotation update)
b Major DNA-binding protein (MDBP).
b U41.
b Human herpesvirus (type 6 / strain Uganda-1102) (HHV6).
b Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
b Betaherpesvirinae; Roseolovirus.
b NCBI_TaxID=10370;
b [1]
b SEQUENCE FROM N.A.
b MEDLINE=95266321; PubMed=7747482;
b Compels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,
b Martin M.E., Efstathiou S., Craxton M., Macaulay H.A.;
b "The DNA sequence of human herpesvirus-6: structure, coding content,
b and genome evolution."
```

Virology 209:29-51(1995).

-!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA REPLICATION.

-!- SUBCELLULAR LOCATION: Nuclear (Probable).

-!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X92436; CAA63167.1; -

EMBL; X83413; CAA58375.1; -

InterPro; IPR000635; Viral DNA bind.

PFam; PF00747; viral DNA bp, 1

KW DNA-binding; DNA replication; Zinc-finger; Nuclear protein.

FT ZN FING 459 475

SQ SEQUENCE 1132 AA; 127762 MW; 4468D3B559F02D8 CRC64;

Query Match 7.8%; Score 492.5; DB 1; Length 1132;

Best Local Similarity 21.8%; Pred. No. 2.3e-29;

Matches 274; Conservative 185; Mismatches 544; Indels 253; Gaps 53;

QY 2 ENTQKTVPTGPGYVYACRVEDLDLEEISFLAARSTDLDALLPLMRNLTVKRTSS 61

Db 3 DENETWSAPVSTAAMIYFPKRELVDLVSLIMERNSPIVISPLMLNLTVENDEFTT 62

QY 62 LAWSGARTTGLAGAGITLKLTTSHFYSVFEVGKHLV-PSAAANLTFACNAARERF 120

Db 63 V-----KTFITPFGGILTKITSFMPVCFPGTQOLVGMAEDHGLRLCEQTQKF 115

QY 121 GFSRCQGPVPGAVETTGAEICTRLGLEPENTILYLVLTALFKEAVFMCNVLHYGLDI 180

Db 116 HLOSPEVETARKVIDIKA--LCSAVGKADSVICHVACGNGFKELF-----AGLI 165

QY 181 -----VHNHGQVIRILPFPVQLFMPDVNRLVPPFNTHRSIGEGFYPTFFNTG 232

Db 166 PCVEEQIOVQGEVSCVKIPLYSATLFETE-----ETISLSSCTEFIOBERGFLPA 216

QY 233 LCHLHDCVIAPMAVALRVNVTAVARGAARLAFDENHEGAVLPDPDITYVFOSSSGTT 292

Db 217 LSETLFYVYFTSWGTLRFSNTKELIDAGLKQFDGEOIVKLAPHKTYL----- 266

QY 293 TARGARNVNSTKPSGSGFERRLASIMAADTALHARVFN-----TGIYEETPTDI 346

Db 267 ---GISGOKISAVEKD-----FLMLVDSV-TELSFSGHVARVYLDVVDV--PSQI 309

QY 347 ---KEWPMFIGEGTLPRLNALGSYVARVAGVIGMVFSPNSALYLTEVE-----D 394

Db 310 MNFNDWFIIRNSETHAERMAQLNLKHLSSHLLAVLIFAPNSILYCKLAFIPNVKQAFN 369

QY 395 SGMTAKDGGPGSFNRYQFAGFHLAANPQTRDGH-VLSSQSTGSSNTEFSVDYLALI 453

Db 370 SVMTQ-----ELLRSLSFCNALSLSDVDVNDNRKLIKCDSTGKDDKFSANLAYA 422

QY 454 CGFGAPILARLLFYLERCDACAGTGGHGDALKVTVTTFDSEIIPCSCERHTRPVCAHTTV 513

Db 423 CATSPQLLSFVWNLNRMSVYNAGNAHTEIYNHLVNC--SALNCEFCDGKCCQSCIGTAM 480

QY 514 HRLRQRMFRFGQATROPIGVFTWMNSQYSCDPLGNVAPYLIIRKPGDQTEAAKATWQDT 573

Db 481 VRVGTRLPAIPKVKKEPLVMSMFSYAYAEVDILGSGF-----RKFSVSELKGIKQDQNT 535

QY 574 YRATLERLFDLEQERLLDRGA-----PCSEGLSSVIVDHPPTFR-----RILPT 618

Db 536 LS-----LDRKEVVSQIFDYCKNSLIDPVTGDTFNVRSKKDFVSIING 580

QY 619 LRARIEQTTQFMKVLVETRDYKI-REGLSEATHSNALTFDPSGAFCEITNPLVKRTHL 677

581 LTQCIECVS---RCIVEMRETQTPRECIENCLQSFVNDVTPYATASPELFTSYKVL 637
DR PIR; T44001; T44001.
DR InterPro; IPR000635; Viral DNA bind.
DR Pfam; PF00747; Viral DNA bp. 1
KW DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
FT ZN_FING 459 475
SQ SEQUENCE 1132 AA; 127817 MW; B39ABAE6A5103EFF CRC64;
Query Match 7.8%; Score 490.5; DB 1; Length 1132;
Best Local Similarity 22.0%; Pred. No. 3.2e-29;
Matches 278; Conservative 180; Mismatches 535; Indels 273; Gaps 54;
QY 2 ENTQKTVTPGPIGVYACRVEDLLEISFIAARSTDSDLLPLMLNLTVEKTFSS 61
DB 3 DENETVVSAPYSTAAWYVFPKKEKELLDVLVSLMERNSPVVISPLMLLTVENDFSTT 62
QY 62 LAVVSGARTTCLAGATLTKLTTSHEVPSYFVHGGKHLV-PSAARNLTACNAARERF 120
DB 63 V-----KTPITNFGTITLTKITSFMPVCFPHGTETQLVGMADHGDLRLCEQTRQKF 115
QY 121 GFSRCQGPVDGAVETTGACICTRLGLEPENTILYLVYVTFALFKEAVFMCNVLHYGLDI 180
DB 116 HLQSFVEPTARKVIDIKA--LCSAVGKADSVICHVACGNGFKELF-----AGLLI 165
QY 181 -----VHLNHGDVIRIDLPVQLEFMPDVNRLVDPFNTHRSIGEGFYVTPFVNTG 232
DB 166 PCVEEQIQVQGEYSCKIPLYSATLFET-----ETLSLSCSTEFIERGFFLPA 216
QY 233 LCHLHDCVIAPIVAVLRVNTAVARGAAHLAFDENHEGAVLPDDITYTYFOSSSGST 292
DB 217 LSETLFYVYVFTSMGTLRFNTKELIDAGLKQFQDGEQVTKLAPHKYL----- 266
QY 293 TARGARNDVNSKSPSGPERRIASIVAAUTALHAEIFN-----TGIVETPTDI 346
DB 267 ---GISGCKISAVEKD-----FLMLVDSVV-TLSFHVAYELDSVYD--PSQI 309
QY 347 ---KEWPMFIMGEGTFLRLNALGSTARVAGVIGAMVSPNSALYLTEVE-----D 394
DB 310 MNENDWPIIENSETHAERMAQLTNLKLHSSHLAVLIPANSILYCSKLAPIFNWQAFN 369
QY 395 SGWTEAKDGGPSPFRNFYOFAGPHLAANPQTRDGH-VLSQSTGSSNTEFSDVYLA 453
DB 370 SVMTQ-----ELLRLSLSFCNALSLTEDVYDNRKIKCDSTSGKDKFSANHLA 422
QY 454 CGFGAPILALLFLYERCDAGFTGGHGDALKVYTCGFDSEI-----PCSLCEKH 503
DB 423 CATCPQLLSVIVNLRMSV-----YNAGNAVTEIYNHLVNSANLCEFCDGK 470
QY 504 TRPVCAHTTVHRLRORMPRFGQATROPIGVGTMSQYSDCDPLGNAPYLIIRKPGDOT 563
DB 471 CCOSCIGTAMVVGTVRLPALPKNVKKEPLVMSMFYSYAEVDILGSGF-----RKEVSEL 525
QY 564 EAKATMODTYRATLERFLFIDLEQERLLDGA-----PCSEGLSSVIVDHPFR--- 613
DB 526 KEIKGQOQNTLS-----LDGRKFVSQIFDYCKNSLIDPVGEDTFNVS 570
QY 614 -----RLDPLRARIQTTFQFMKVLVETRDYKI-REGLSEATHSMALTDPYSGAFPI 667
DB 571 KDOFVSIIHGLTQICECVS---RCIVEMRETQTPRECIENCLQSFVNDVTPYATASPE 627
QY 668 TNFLVKRTHLAVVQDIALSOCHCVFVGQOQVEGRNFNFQPVLRFRFVDLF-----N 719
DB 628 LTFAYYKVLTVLQNLAL-----IVASGHVVDPRCTGNSISKMLVQOQYSLYGTFFHSSYLK 683
QY 720 GGISTRISIVTLSEGEVSAPEPTLQ--DA---PAGR-----TFDGLARVSVEVIRDIR 770
DB 684 KGFENRTVKV-----ASNVDMEQILDCLYSGKVKYKTTIOAKLRLSNQCLURDF 735
QY 771 VKNRVFGNCTNLSEARARLVGLASAYQREKRVDMHGLGFLKQFGLLFRGMP 830
DB 736 IKRNPFSKSKTAHNNPYFKNV-----KHKKNPLSGCISFLPKYHDKLF----- 781
QY 831 PNSK-SPNPQWFVTLQRMQPADKLTHEEIT-IAAVKRPTEEYAAINFLPTCIGE 888

581 LTQCIECVS---RCIVEMRETQTPRECIENCLQSFVNDVTPYATASPELFTSYKVL 637
678 AVVQDLALSOCHCVFYQOQVEGRNFNFQPVLRFRFVDLF-----NGGFSTRSIT 729
638 TVLQNLAL-----IVASGHVVDPRCTGNSISKMLVQOQYSLYGTFFHSSYLKQFGLNTRTVK 693
730 VTLSEGPVSAPEPTLQ--DA---PAGR-----TFDGLARVSVEVIRDIRVKNRVFSGN 780
694 V-----ASNVDMEQILDCLYSGKVKYKTTIOAKLRLSNQCLURDFRIKRPNKSS 745
781 CTNISEARARLVGLASAYQREKRVDMHGLGFLKQFGLLFRGMPNPKS-SPNPQ 839
746 KTAHNNPYFKNV-----KHKKNPLSGCISFLPKYHDKLF-----PNVKISLLEL 791
840 WFWTLQRMQPADKLTHEEIT-IAAVKRPTEEYAAINFLPTCIGELAQFYANLI 898
792 WQRFLLNNVKTLDIGNPEEVKTFIKAFSITNTYDEIDIDIQECLSTFIDCYFHNKF 851
899 LKYCHSQYL--INTLTS-IITG-----ARRPRDS--SVLHWIKDVTSAADTET 944
852 LSALGFHDYLTSLHGLTSKLVQNPVLPFVVDLQKPFSSIQEYLYVYKLVLDG--VFN 909
945 QAKALLEKTENLPELWTATSTHLVRAAMNQPMVVLGISISKYHGAAGNNRVFOAG-- 1002
910 PVIASLSKEPN---FGTIPTS-----RSLVTFGLTLEKPVSLA-NREYFQFQL 954
1003 NWSGLNG-GRNVCP-----FTFDRTRRFLIACPRGGRICPVTPGSSGNRETLSDQ-- 1053
955 GWIGSGVDRLNAPLTSALQDFMR--OKTIIIAIKFSEVIV-----KVRRAIMEDTSEV 1008
1054 VRGIIVS-----GGAMVQIAIVATVVRVAGAAHMAFDMLSLTD--EFLARD----L 1102
1009 VKGKVLSEVENLTNDIDPELLIAEVMR-----DREDKPTMDMLFFVDGREALAASIMLKL 1065
1103 EELHDOIQIULEPTWTEGAEAV-----KILDEKTAGDG-----ETPT 1142
1066 NHLVDVNVKQFSIT-NLQSFVETSSNDAPVYDFSEILAEEDDQGVNLKCDDETET 1120
RESULT 13
DNBI HSV62
ID DNBI HSV62 STANDARD; PRT; 1132 AA.
AC P52538;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major DNA-binding protein (MDBP).
EN U41 OR XA2L.
OS Human herpesvirus (type 6 / strain Z29) (HHV6).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Roseolovirus.
OX NCBI_TaxID=36351;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95074921; PubMed=7983761;
RA Stamey F.R., Dominguez G., Black J.B., Danbaugh T.R., Pellett P.E.;
RT "Intragenomic linear amplification of human herpesvirus 6B orilyt
RT suggests acquisition of orilyt by transposition.";
RL J. Virol. 69:589-596(1995).
SC -|- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
SC REPLICATION.
SC -|- SUBCELLULAR LOCATION: Nuclear (Probable).
SC -|- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
SC FAMILY.
SC -----
SC This SWISS-PROT entry is copyright. It is produced through a collaboration
SC between the Swiss Institute of Bioinformatics and the EMBL outstation -
SC the European Bioinformatics Institute. There are no restrictions on its
SC use by non-profit institutions as long as its content is in no way
SC modified and this statement is not removed. Usage by and for commercial
SC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
SC or send an email to license@isb-sib.ch).
SC -----

782 PNVKISLELMQRELLNNVPTKIDIGNPEVVKTFIKFAFSITNTYDEIDDIIDIOPECLST 841
889 LAQYMANLILKYCDHESYL--INTLTS-IITG-----ARRRDPSS--SVLHWIRK 934
842 FIDCYFANKLSALGFHDYITSLHGLTSKLVTONVLPVVLQKPKFSSIQBYLYVVK 901
935 DVTSADIETQAKALEKTENPELMTAFTSTHLVRAANNQPMVVLGISISKYHGAAG 994
902 LVLDG--VNPVPIASLSKEPN---FOTIITS-----RSLVTFGLTLEKFSVSLA- 944
995 NNRVFOAG--NWSLNG-GRNVCP-----FPTDTRRPIIACPRGGICPVTGPSSGN 1045
945 NREYFPGQUGWIGGSVDNRNLPTNSALODFRFMR-KQTIITATKFSIV-----KKVR 998
1046 RETLSDQ--VRGIIVS-----GGAMVQIAIVATVAVRAVGAQAQMAFDDMLSLTDD-E 1096
999 RETIMFTEVVKVSLIVENLNDIDPELLIIAEVWR---DREDKPTWDDMLFFVDGRE 1055
1097 FLARD-----LEELHDQIIQLETPWTVEGALEAV-----KILDEKTTAGDG--- 1138
1056 ALAASIMLNHLNLMVNRVDFSA-NLQSVFEAVSSNDAPVYDFSEILAEEDDQASGVLK 1114
1139 --ETPT 1142
1115 CDETET 1120

RESULT 14
NBI_HSV7J
D DNBI_HSV7J STANDARD; PRT; 1131 AA.
T 01-OCT-1996 (Rel. 34, Created)
T 01-OCT-1996 (Rel. 34, Last sequence update)
T 16-OCT-2001 (Rel. 40, Last annotation update)
E Major DNA-binding protein (MDBP).
N U41.
S Human herpesvirus (type 7 / strain J1) (HHV7).
X Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
X Alphaherpesvirinae; Simplexvirus.
X NCBI_Taxid=52728;
N [1]
N SEQUENCE FROM N.A.
P STRAIN=J1;
L Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
L -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
C REPLICATION.
C -!- SUBCELLULAR LOCATION: Nuclear (Probable).
C -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
C FAMILY.
C
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (see <http://www.isb-sib.ch>)
C or send an email to license@isb-sib.ch.
C
R EMBL: U43400; AAC54703.1; -
R PIR: T41943; T41943.
R InterPro: IPR000635; Viral DNA bind.
R Pfam: PF00747; Viral DNA bp; 1.
R DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
T ZN_FING 459 475 C2HC-TYPE
Q SEQUENCE 1131 AA; 129008 MW; 52697388D1B6D04F CRC64;
Query Match 7.7%; Score 483; DB 1; Length 1131;
Best Local Similarity 20.8%; Pred. No. 1.2e-28;
Matches 258; Conservative 204; Mismatches 560; Indels 220; Gaps 46;
2 ENTQKTVTPGPIGVYVACHVEVDLDEESFLAARSTDSDLLPLMRNLTVKFTSS 61

Db 3 DNETVVSAPICTAAWYILPKQKLEILTTLSEMKRKSWSIPLLNLTVENDEPFT 62
QY 62 LAVVSGARTTGLAGAGITLKLTTSHFVPSVFRHGGKGLVPSA-APNLTRACNAARERF 120
Db 63 V-----KTFIIVGGTVITKITSFNPVCFVFFHGTVDVFLKEADHGNLKLCKQTRKF 115
QY 121 GFSRCQGPVGVAVETTG-ABICTRGLGEPENTILYLVTTALPKEAVFMCNVFLHYGLD 179
Db 116 NL-----QEFVNGNRKSDIGKICESVGRNADDVLCHIVVNGNGFKELF-----AGLL 164
QY 180 I-----VHNHGD--VIRIPLFPVQLFMPDNLVDPDPENTHRSIGEGFVPTPEYNT 231
Db 165 IPCVEEQIQVQGECLAIKIPLYSATLFESEELCI-----DTCTEFIGENFYAP 215
QY 232 GLCHLIHDCVIAPMAVALRVNNTAVARGAAHAFDENHEGAVLPDITVTYTPQSSSGT 291
Db 216 QISEVLFYLIFTSGWMTLRFNNTLELIKAGLKQFIQDTEQTVKLA PNKY----- 265
QY 292 TTARGARRNDVNSTKPSGSGGFERRLASIMAAIDTALHAEVFN-----TGIYEETP-T 344
Db 266 --HGIPGQKLSPIED-----HMLVLD-AVITELTFSYTAEYLDSDIYENQIM 310
QY 345 DIKEMPMFIGMEGLPLNALGSIYARVAGVIGAMVFPSPNSALYLTVEVDSGTEAKDG 404
Db 311 NFSEWPIIKSAETHEEKIVELKXLRHLSHVAALVFAANSILYSNKLAYISNWK----- 365
QY 405 PGPSFN-----RFYQAGPHLAANQTDGDH-VLSQSTGSSNTEFSDVYLALIC 454
Db 366 --QAFNSAITQETLLRSIOFCNSLSINEDFYNDARKLIKNSSPCKEDKFAFLAYAC 423
QY 455 GFGAPLLARLLFYLERCDAAGATGGHGDALKVGTGTFDSEIPECSICEKHTRPVCAHTVH 514
Db 424 ATCPQILSHIWNLSNINCGNSEIYNHIVNC--SSNLCEFCGCKCHSCITGTAII 481
QY 515 RLQRMPRFGQATRQPIGVGTWNSQYSDCPGLNYPAYLILRKPGDQTEAAKATMDTY 574
Db 482 RINSRLPQISKTTKEPIVMTMFSRFYADVDLVSGFGKGVNESKDPKMEAQTPTSLDRP 541
QY 575 RATLERLFIDLEQERLLDRGAPCSSEGLSSVIV--DHPTFR-----RLDTRLARIQT 626
Db 542 K-----FLGMHDY-----CKKNLIDALTGEDNLFKSNQDNFVNMINDLIQIEEA 588
QY 627 TQPMKVLVETRDYKIREGLSEATHSMALTDPYSGAFCPITNFLVKRTHLAVVDLALS 686
Db 589 VSKICSEMEKTO--TSREQIENCLOSFNIDTTPLSAFSPFPVTVYKVVILVQLNAL- 645
QY 687 QCHCVFYGOVE-----GRNFRNQPVLLRRFVLDLFGGPFSTRSITVTLSSEGPS 738
Db 646 ---IIGTVVDPRDCTGNLISKWLMQOYQSLYGAFYNGHFKGFLNMKTVKI----- 694
QY 739 APNPTLQ-----DAPAGRTPDGLARVSVEIRDIVKRVVFSNGCTNLSEAR 789
Db 695 ASNVDMEQVIDFNLFKSGKYAKTSIQAKLCLSLWQCLXDFVKRPPFNKNTQNN--- 751
QY 790 ARLVGLASAYQOEKRVDMHGLQGLFKQPHGLLFPFGMPNPSKSNPQPFWTLQRNQ 849
Db 752 -----PFFKVKQKQNPGLSGCLSLFLFKYHERLF-----PNLKISCLE-FWQRIILNN 798
QY 850 MP--ADKLUTHEIIT-IAAVKRFTEEYAAINFNPPTICIGELAQFYMANLILKYCDHSQ 906
Db 799 MEKTDIGNVEDMRSPFKFTFRVTVNSYDEIDLDIQECLLSFIETTFHNKLLSVLGYRD 858
QY 907 YL--INTVTSITGARRPRDPSSVLHWRKQVNT--SAADIEIQAKALLEKTENPELWTT 962
Db 859 YLTSLHALTSKLIV-----PQNDMLFPVFLKEHPTFSSVQEVVHVHVKLV--GNGLKEPMTA 912
QY 963 APTSTHLVRAANNQPMVVLGISISKYHGAAGNN--RVFQAGNWSGLNGKNTVCLP----- 1016
Db 913 SLTKEPNFGSFTGSIITFGIMIEKFPVSVASRDYFHFQGLGWTAGSGVDRNLNPPSSGL 972
QY 1017 --FTPDRTRFIIACPRGGFICPVTGPSSGNRETTLSD--QVRGII-----VSGGAMVQ 1066

Db 973 QDFRMR-QKFVIATK-...LCDII-VKKVKEATVDEVIRGKVLNIIESLSNVPE 1026

QY 1067 LAIYATVVRVAGARQAQMAFDMLSLTD-BELARDLE-...ELHDOIQTLETP 1116

Db 1027 LLILAEMK-...DRDSKTMDDMLFYVDRPLAKSVNKKIQTHTDNLNVHDFSLSTLLS 1083

QY 1117 W-...TVEGALEAVKILDEKTTAGDEPT 1142

Db 1084 FEEQVEDSAAYDFSELLVEGNEQFGILKCEETEHENEEPS 1125

RESULT 15

DNBI_SCMVC

ID DNBI_SCMVC STANDARD; PRT; 1160 AA.

AC P13215;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Major DNA-binding protein (MDEP).

GN UL27 OR DBP.

OS Simian cytomegalovirus (strain Colburn).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Betaherpesvirinae; Cytomegalovirus.

OX NCBI_TaxID=50292;

RN [1]

RX MEDLINE=91037979; PubMed=2172458;

RA Anders D.G.;

RT "Nucleotide sequence of a cytomegalovirus single-stranded DNA-binding protein gene: comparison with alpha- and gammaherpesvirus counterparts reveals conserved segments.";

RT J. Gen. Virol. 71:2451-2456 (1990).

RL [2]

RN SEQUENCE OF 205-308 FROM N.A.

RP MEDLINE=88155776; PubMed=2831398;

RX Anders D.G., Gibson W.;

RA "Location, transcript analysis, and partial nucleotide sequence of the cytomegalovirus gene encoding an early DNA-binding protein with similarities to ICp8 of herpes simplex virus type 1.";

RL J. Virol. 62:1364-1372 (1988).

CC -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA REPLICATION.

CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).

CC -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR ENBL; D00750; BAA00647.1; -.

DR ENBL; M19868; AAA46066.1; ALT SEQ.

DR InterPro; IPR000635; Viral DNA bind.

DR Pfam; PF00747; Viral DNA bp. 1.

CC DNA-binding; DNA replication; Zinc-finger; Nuclear protein;

CC Early protein.

CC ZN FING 465 479 C2HC-TYPE.

CC SEQUENCE 1160 AA; 129005 MW; 7037716816974B1A CRC64;

CC

Query Match 7.6%; Score 479.5; DB 1; Length 1160;

Best Local Similarity 21.3%; Pred. No. 2,3e-28;

Matches 279; Conservative 202; Mismatches 57; Indels 255; Gaps 53;

Y 1 MENTQKTVTVTPGPGYVYACVREDLDEISFLAARSTDSLALLPLMRNLTVKFTFS 60

Y 1 MGNEISALAPYGPAAVYVFTKTHMNEVLATSLCSDSSPVVIAPLMGLTVDDQFCT 60

Y 61 SNAVSGARTTGLAGAGITLKTTHSHYFVSVFHGGKHVLP-SAAAPNLTRACNAARER 119

Db 61 SV-...RTPVVCVGGVLTWKVS-PCFFALYFNTQGVDFSEPHGDVQRLCDETRQR 113

QY 120 GFGRSQQPPVDGAVETTGAEICTRLGLEPENTILYLVVLTALFKEAVFMCVNLHYGLD 179

Db 114 YAESVM-PEEGRAPTDLAALCTAAGCDPOBVLVHVVGNGMEFMYAGOLICFPEAA 171

QY 180 IVHINHGIVIRIPLFPVOLP--MPDVNRLVDPDPNTHRSIGEGFVYPTFYNTGLCHL 236

Db 172 PTRLNDCDAVRVLPYPTLFGSLQADV--SDELSLDKES--SFVESRGLYVPAYSET 225

QY 237 IHDCVIAPMVALVRNVTVAVARGAAHLAFDENHEGAVLPDITITYFQSSSSQTTTARG 296

Db 226 LFYVYVTSWCQALRFSETKVLEAALKQFVNDSSQSVKLAPHKKYFGYTSOK----- 277

QY 297 ARRDVNSTKPSGSGPERRLASIMAADTAL-----HAEVINTGIYBETPTDIKWP 350

Db 278 -----LSSLEK-----HMLSDAIVICLGFSGFASVFLDSAYGASDSMVYSEWP 321

QY 351 MFIGMEGTLPRINALGSYARVAGVIGAMVPSNSAL-----YLTEVEDSGMTEAKDGG 404

Db 322 VVWATDHRDLIRALTELKHLSTHISALLFSCNLSILYHNRLLVLTENKVASGT----GA 377

QY 405 PGPSFNRYQFAGPHLAANPQTD RDGH-VLSSQSTGSSNTEFSVDYLALICGFCAPLLAR 463

Db 378 SQEVLKSIHFANGLTGLCEDTYNDARKLKCSGWAKDERYAPYHLSLICGTCPLFSA 437

QY 464 LLEFLERCDAGFTGCHGDALKYVTGTFDSE-----IPC--SLCEK-----HTRPVCA 509

Db 438 FIWTLNRVSV-----YNTGLTGSSTLSNHLIGCSSLGCACGCTCCT--CY 482

QY 510 HTVHRLRQRMRFQGAQTRQPIGVFGTMSQYSDCDPLGNAPYAPYLILRKPQDQTEA--- 565

Db 483 NTAFRVQTRLPQMPRLPKKPSVVMQSRFLNDVDVLGTFG-----RRYSAESKEASLDA 538

QY 566 -----AKATMQTYRATLERLIDLEQERLLDRGAPCSSEGLSSVVD-HPT 611

Db 539 KADGSGASTNRTASSSVDRTHR--LNRILDYCKQWELID-----SVTGEDTNTINGRSD 591

QY 612 FRRILDTRARIEQTTFQFMKVLVETRDYKIREGLSEATHSMALTDPYSGAFCPITNEL 671

Db 592 FINLVSSLNKVFDDKAMS FVS---EVRMKSNDREVLGATQAFNLDLNPFAVSFPLAVE 648

QY 672 VKRTHLAVODLALSQCHQVYGC-----QVEGNFRNQFPVLRFRFVDFLNGGFISTRS 727

Db 649 YYRVIFAIIONVALITATSYVDNPLTTSLSVRWVTHQFQSI-----HGAFSTTS 599

QY 728 ITVTLSEGPVSAENPLTGDAPAGR-----TFDGLARVSVVEVIRDIRVK 772

Db 700 -----RKGLFIRNVKSKNADHRLPDFKLVARGYTSVISMEIKLSRLSVPSLLMFRVK 754

QY 773 NRVVFGNCTNLSEAAARLVGLASAYORQE--KRVDMHLGALGFLKQFHGLLPRGMP 830

Db 755 NRPI-----SKASKGT--TAHVFFRRRHVPKCNPKVKGCLGFLLYKYHKLFP---- 799

QY 831 PMSKSPNQMFVTLQRONQPADKLTHEITIAVKRF-----TEEYAAININLPP 883

Db 800 --DCGFSCLQFQKVCANALP-----KNVNIQMGFFNFVKEFVISVTADYNHEDLIDVPP 853

QY 884 TCIQELAQFYMANLILKYCDHSQYL--INTLTSIITGARRPRDPSSVLHWIRKQDVTSAAD 941

Db 854 DQMLNYLENRPHNKFLCFYGFQDYIGTLHGLTTLTYQNHAQFP--YLLGESPNFASAD 911

QY 942 IETQAKALLEKENIPELMTTFTTHLVRAAMNQRPVVLGIS:SKYHGAAGNNRVFQA 1001

Db 912 FALRLKDL--KATGVTAPLASTVTRESLMRTIFEQRLSVTSFSTIEKYAGVNNKKEIYQF 969

QY 1002 GN-----WSG-----LN-----GGKNVCPLFTDPTDRERFIACPRGGFCFVTCPSGNET 1048

Db 970 GQIGYFSGNGVRSINTNSIGQD-----YKFWR-ORCLATKLSDLVI-----KSRSDN 1019

QY 1049 TISDQVRGIIVSGGAMVOL-----AIYATVVAVGARAQHMVAFDDWLSLTD-- 1095

Db 1020 VLFDE---DIKRWAAALSDENLDVDPMLMAY-----EILSTREEIPERDDVLFVDGC 1072

1096 EFLARDLEELHDQIIQTLTETPTWVEGALAVKILDEKTTAGDGETPTNLAEPDSCPSH 1155
 1073 QAVADSLMEKFSRLOEMGVDDFSL---VNLQQVLDSRPECGGGGEV-----H 1117
 1156 DTTSNVLNIGSNISGTVPLK-RPPED--DELFDLSGIFIKHNI 1199
 1118 DLSALFTAASGEAVGNSV--GLNARGGSHAFDE--DCOLLPAKRGRL 1160

Search completed: January 30, 2004, 13:13:04
 OS time : 22 secs

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

MM protein - protein search, using sw model

run on: January 30, 2004, 13:10:47 ; Search time 27 Seconds
(without alignments)
4284.849 Million cell updates/sec

Title: US-09-769-699-2

Effect score: 6294

Sequence: 1 MENTQKTVTPTGPIGVYA.....DEFLDGLPIKHNITWEM 1203

Coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR 76.*

- 1: PIR1.*
- 2: PIR2.*
- 3: PIR3.*
- 4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6294	100.0	1204	1 DNBEC29	DNA-binding protei
2	3620.5	57.5	1209	1 DNBEC4	DNA-binding protei
3	3583	56.9	1208	2 T42574	DNA-binding protei
4	3054	48.5	1196	1 DNBKS	DNA-binding protei
5	3049	48.4	1196	1 DNBK1	major DNA-binding
6	3047	48.4	1196	1 DNBK1	DNA-binding protei
7	3003	47.7	1197	1 A48350	DNA-binding protei
8	2933	46.6	1186	1 DNBEG	DNA-binding protei
9	1094.5	17.4	375	2 JQ0846	DNA-binding protei
10	669	10.6	1128	1 Q0B847	major single-stran
11	633	10.1	1128	2 T42922	DNA-binding protei
12	618	9.8	1145	2 S55600	single-stranded DN
13	604	9.6	1128	1 DNBEM1	DNA-binding protei
14	596	9.5	1127	2 T03105	major single-stran
15	490.5	7.8	1132	2 T44001	major DNA-binding
16	483	7.7	1131	2 T41943	major DNA-binding
17	479.5	7.6	1160	2 A36256	DNA-binding protei
18	477	7.6	1235	1 Q0BEM4	DNA-binding protei
19	412	6.5	1191	1 A44051	DNA-binding protei
20	237.5	3.8	483	2 S69894	major DNA-binding
21	137	2.2	1504	2 T17426	FK506 polyketide s
22	120.5	1.9	781	2 A13014	methyl-accepting s
23	120.5	1.9	788	2 F98269	hypothetical prote
24	119.5	1.9	701	2 A11501	internalin protein
25	117.5	1.9	358	2 H83380	O6-methylguanine-D
26	116.5	1.9	335	2 H81702	adherence factor T
27	115.5	1.8	756	2 T17428	FK506 polyketide s
28	115	1.8	823	2 B35963	protein-tyrosine k
29	115	1.8	2505	1 XYRTFA	enoyl-[acyl-carrie

ALIGNMENTS

RESULT 1

DNBE29

DNA-binding protein - human herpesvirus 3

C:Species: human herpesvirus 3, varicella-zoster virus

C:Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999

C:Accession: C27214

R:Davidson, A. J., Scott, J. G., & E. W.

J. Gen. Virol. 67, 1759-1816, 1986

A:Title: The complete DNA sequence of varicella-zoster virus.

A:Reference number: A27345; MUID:86306657; PMID:3018124

A:Accession: C27214

A:Molecule type: DNA

A:Residues: 1-1204 <DAV>

C:Cross-references: EMBL:X04370; NID:G59989; PIDN:CAA27912.1; PID:G60018

C:Genetics:

C:Gene: 29

C:Superfamily: herpesvirus DNA-binding protein

C:Keywords: DNA binding

Query Match	100.0%;	Score 6294;	DB 1;	Length 1204;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1203;	Conservative	0;	Mismatches	0; Indels 0; Gaps 0
QY	1	MENTQKTVTPTGPIGVYACRVEDLLEISFLAARSTDS	LALLPLMNLTVKFTFS	60
DB	1	MENTQKTVTPTGPIGVYACRVEDLLEISFLAARSTDS	LALLPLMNLTVKFTFS	60
QY	61	SLAVVSGARTTGLAGAGITKLTTSHFYPSVVFHGGKHL	PSSAAPNLTRACNAARERF	120
DB	61	SLAVVSGARTTGLAGAGITKLTTSHFYPSVVFHGGKHL	PSSAAPNLTRACNAARERF	120
QY	121	GFSCQGPVVDGAVETTGAEICTRLGLEPENTILYLVTAL	FKEAVFMCNVFLHYGGGLDI	180
DB	121	GFSCQGPVVDGAVETTGAEICTRLGLEPENTILYLVTAL	FKEAVFMCNVFLHYGGGLDI	180
QY	181	VHINHGDIVIRIPFPVOLFMPDVNLVPDPENTHRSIG	BGVYPTFYNTGLCHLHDC	240
DB	181	VHINHGDIVIRIPFPVOLFMPDVNLVPDPENTHRSIG	BGVYPTFYNTGLCHLHDC	240
QY	241	VIAPMAVALRVNVTAVARGAHLAFDENHEGAVLPDIT	TYTFQSSSGTTTARGARN	300
DB	241	VIAPMAVALRVNVTAVARGAHLAFDENHEGAVLPDIT	TYTFQSSSGTTTARGARN	300
QY	301	DVNSTSKPSGSGFERELASIMAAOTALHAELVNTGIVYEET	PTDIKEWPMFTIGEGTLP	360
DB	301	DVNSTSKPSGSGFERELASIMAAOTALHAELVNTGIVYEET	PTDIKEWPMFTIGEGTLP	360
QY	361	RLNALGSYTTARVAGVICAMVFPNSALYTEVEDSGMTAKD	GGPGPSFNRFYQFAGPHL	420
DB	361	RLNALGSYTTARVAGVICAMVFPNSALYTEVEDSGMTAKD	GGPGPSFNRFYQFAGPHL	420
QY	421	AANPQTRDRGHVLSQSSNTSFSDVILALICFGFAGPILARLL	FYLERCDAAGFTGGH	480

09/769,699
2-21-04
Search Notes

```

Db 421 AANPQTDGSHVLSQSTGSSNTEPSVDYLALICGFGAPLARLLFYLERCDAGFTGGH 480
QY 481 GDALKYVGTGDSSEIPCSLCEKHTRPVCAHTTVHRLRQMRPFQOATRPQIGVGTNSQ 540
Db 481 GDALKYVGTGDSSEIPCSLCEKHTRPVCAHTTVHRLRQMRPFQOATRPQIGVGTNSQ 540
QY 541 YSDCDPLGNVAPYLLTRKPGDQTEAAKATMDQTYRATLERLFDLEORLLDRGAPCSSE 600
Db 541 YSDCDPLGNVAPYLLTRKPGDQTEAAKATMDQTYRATLERLFDLEORLLDRGAPCSSE 600
QY 601 GLSSVVDHPTFRILDLTLRARIISQTTQFMKVLVETRDYKIREGLSEATHSMALTFDPY 660
Db 601 GLSSVVDHPTFRILDLTLRARIISQTTQFMKVLVETRDYKIREGLSEATHSMALTFDPY 660
QY 661 SGAFCPITNFIKVLTHLAVODLALSOCHCVFQGVQOQVEGRNFRNQFVLRPFVDLFG 720
Db 661 SGAFCPITNFIKVLTHLAVODLALSOCHCVFQGVQOQVEGRNFRNQFVLRPFVDLFG 720
QY 721 GFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTPDGLARVSVEVIRDKVRNVFSGN 780
Db 721 GFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTPDGLARVSVEVIRDKVRNVFSGN 780
QY 781 CTNLSEARARLVGLASAYQOEKRVDMHLGALFLKQPHGLLFRGMPNPKSPNPQW 840
Db 781 CTNLSEARARLVGLASAYQOEKRVDMHLGALFLKQPHGLLFRGMPNPKSPNPQW 840
QY 841 FWTLLQRNOMPADKLTHEEITTAIVKRFTEEVAAINFILNPTCIGELAQFYMANLILK 900
Db 841 FWTLLQRNOMPADKLTHEEITTAIVKRFTEEVAAINFILNPTCIGELAQFYMANLILK 900
QY 901 YCHSQVILNLTLSITGARRPDPSVLHWIKDVTSAAIDTQAKALLEKTENLPELW 960
Db 901 YCHSQVILNLTLSITGARRPDPSVLHWIKDVTSAAIDTQAKALLEKTENLPELW 960
QY 961 TTAFTSTHLVRAAMNORPMVLGISIKVHGAGNNRVFOAGNWSGLNGKNCVCLFTPD 1020
Db 961 TTAFTSTHLVRAAMNORPMVLGISIKVHGAGNNRVFOAGNWSGLNGKNCVCLFTPD 1020
QY 1021 RTRRFIIACPRGFCFVTPGSSGNRETTLSQDVRGIIYSGGAMVOLAIYATVRAVGAR 1080
Db 1021 RTRRFIIACPRGFCFVTPGSSGNRETTLSQDVRGIIYSGGAMVOLAIYATVRAVGAR 1080
QY 1081 AQHMAFDDMLSDDEFLARDEELHDQIIOTLETPTWVEGALBAVKILDEKTTAGDGET 1140
Db 1081 AQHMAFDDMLSDDEFLARDEELHDQIIOTLETPTWVEGALBAVKILDEKTTAGDGET 1140
QY 1141 PTNLAFNFDSCPSHDTTNSVLNLSGNSISGTVPGKRPDDDELFDLSGIPIKHGNIT 1200
Db 1141 PTNLAFNFDSCPSHDTTNSVLNLSGNSISGTVPGKRPDDDELFDLSGIPIKHGNIT 1200
QY 1201 MEM 1203
Db 1201 MEM 1203

```

```

RESULT 2
NBECA
NA-binding protein - equine herpesvirus 1 (strain Ab4p)
Species: equine herpesvirus 1
Note: host Equus caballus (domestic horse)
Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 16-Jul-1999
Accession: E36798
Submitted to GenBank, March 1992
Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
Description: The DNA sequence of equine herpesvirus-1.
Reference number: A36805
Accession: E36798
Molecule type: DNA
Residues: 1-1209 <TEL>
Cross-references: GB:M86664; NID:G330791; PIDN:AAB02466.1; PID:G330823
Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
irology 189, 304-316, 1992

```

A;Title: The DNA sequence of equine herpesvirus-1.
A;Reference number: A41831; MUID:92295566; PMID:1318606
A;Contents: annotation; possible protein-coding frames
A;Note: neither amino acid nor nucleotide sequence is given
C;Genetics:
A;Gene: 31
C;Superfamily: herpesvirus DNA-binding protein
C;Keywords: DNA binding

```

Query Match 57.5%; Score 3620.5; DB 1; Length 1209;
Best Local Similarity 57.1%; Pred No.1e-266;
Matches 698; Conservative 191; Mismatches 299; Indels 35; Gaps 14;

QY 1 MENTOKTVVTPGPGYVYACRVEDDLLEEISFLAARSTDSDLALLPLMRNLIVKFTFTS 60
Db 1 MESAPKTVSLPVSPLGYVYARQKASLQGTGTVSLTAARSVDSDLAVLVIRGLTVEQFTT 60
QY 61 SLAVVSGARTTGLAGAGITLKLTTSHFVPSVFPVPHGKHVLPSSAANLTRACNAARERF 120
Db 61 NVAVVAGSKTTLGLGTGTLTTPSHFNPAFVYGGSVTGASSNAFNLTRACNAARERF 120
QY 121 QFSRCQPPVDGAVETTGAETCTRLGLEPENTIIYLVVTFALKEAVFMCNVFLHYGLDI 180
Db 121 GFSAFSSPPVENAVETSGEEICASLNLSPETTALYLVVTFSEFKEMVYVCTFLHYGSTT 180
QY 181 VHINHGDIPIPLFPVOLFMVDNRLVDPFNVHHRISIGEGFVYPTPTNTGLCHLHDC 240
Db 181 VTIDQDAMKPIPIPVQLYMPDNVRLASEPFAKHSIGDEFVYSRPFNSDLCLRLHY 240
QY 241 VIAPMAVALRVNVTAVARGAAHLAFDENHEGAVLPDITVTVTFQSS--SSGTTTARGAR 298
Db 241 VLGPAAVALRVNLDVGARGAHLALDENHEGSLVPODVITFTLFDSTQGNAGKSGRAQR 300
QY 299 RNDVNSTKSPSGCFERRRLASINAAADTALHAENVITWGTIYESTPTDIKEWPMFIGMEGT 358
Db 301 QGD--GSGSKNSASGIERRLASVAAADTALSVDISIMGAGIYDTLPSEVDPVLSGDDT 359
QY 359 LPRNLALGSTRVAVGVIGAMVSPNSALVLTVEVDSGMTAKDGGGFSFNRFFYQFAGP 418
Db 360 -ESLEALCAVAAKSLGLVGMVANSVLTWEDDGGADGKGS-NPSYHRFYLIAAP 417
QY 419 HLAANPOTDRGHVL-----SSQSTGSSNTEPSVDYLALICGFGAPLARLLFYLERCD 472
Db 418 YVAGNPOTDRGVLPHTADQQAAPINGSNOBESLDYLALACGFCQILARLLFYLERCD 477
QY 473 AGAFTG-GHGDAKYVGTGDSSEIPCSLCEKHTRPVCAHTTVHRLRQMRPFQOATRPQI 531
Db 478 AGTGGNETDAURYLANLTLESVPCGLCNOATRPACATTLHRLRQRLPRFGAPVRAPI 537
QY 532 GVFGTWNQVSDCDPLGNVAPYLLTRKPGDQTEAAKATMDQTYRATLERLFDLSQERLL 591
Db 538 GIFGTWNQVSDCDPLGNVAPYLLTRKPGDQTEAAKATMDQTYRATLERLFDLSQERLL 596
QY 592 DR-----GAPCSSEGLSSVIVDHPTRRLDLTLRARIISQTTQFMKVLVETRDYKIREGL 646
Db 597 DKETLAQASPCAP--TSVVDHQAQSGFILLNKTIEGAARQFMRILTVEARDFKIREGL 654
QY 647 SEATHSMALTFDPYSGAPCPIITNFIKVLTHLAVODLALSOCHCVFQGVQOQVEGRNFRNQF 706
Db 655 ADANHTMSLSLDPYSSSFCPVTSFLARRTVFAVLQDLVLSQCHCLFYQSGVEGRNFRNQF 714
QY 707 QPVLRFRFVDFLNGGFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTPDGLARVSVEVIR 766
Db 715 QPVLRFRFVDFLNGGFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTPDGLARVSVEVIR 774
QY 767 RDIKVRNVFVSGNCTNLSEARARLVGLASAYQOEKRVDMHLGALFLKQPHGLLFRGMP 826
Db 775 RDLVRNVFVSGNCTNLSEARARLVGLASAYQOEKRVDMHLGALFLKQPHGLLFRGMP 834
QY 827 RGMPPNPKSPNPQWFWTLQRNOMPADKLTHEEITTAIVKRFTEEVAAINFILNPTCIGEL 886
Db 835 RGHPPGIDTNPQWFWTLQRNOMPADKLTHEEITTAIVKRFTEEVAAINFILNPTCIGEL 894

```

```

QY 887 GELAQFYMANLILKYCDHSOYLINTLTSITGARRPRDPSSVLHWIRKDVTSAADETQA 946
Db 895 GELAQFYANLVLYKCDHSQYFINGLTAVWGSRPRDPAVLAWTDRIINGAADVEPAA 954
QY 947 KALLEXTENLPELWTTAFTSTHVRAMQORPMVLGISIKYHGAAGNNRVFOAGNWSG 1006
Db 955 QEVQLQGLGNPAATGTFSTNNRVYMPQRPVWVIGLSIKYNGSAGNNRVFOAGNWSG 1014
QY 1007 LGGKQNVCLPFTDRTRRTIACPRGGFICPVTPGSSGNRETTLSDOVRGIIVSGAMVQ 1066
Db 1015 LGGKQNVCLPMAFDRTRRVFLACPRVGFCEAGGFTGVRENTLSEQVRGIVSEGGPMVQ 1074
QY 1067 LAIYATVAVGARAQAHAFDOWLSLTDDEFLARDELDHDIQTLETPTWVEGALEAV 1126
Db 1075 TAVFAVLHALGARTOHLAVDDMIGLVDDDEFLAASLDALNATVWQF-GEWSVEAAQELV 1133
QY 1127 KILDEKTTAG---DGETPTNLAFFDSC---EPSHDTTNSVLNISGNSISTVPGKLRP 1180
Db 1134 KMEAAQTITAGVAAEGE---AFDFGACVGDTPQOSTSA---FNGGLAMAAAPAGQKRS 1185
QY 1181 PEDDELFDLSGIPKIKHGNITMEM 1203
Db 1186 LPDDILFDMGAPPEKKSGLTFTDM 1208

RESULT 3
42574
NA-binding protein - equine herpesvirus 4 (strain NS80567)
;Species: equine herpesvirus 4
;Variety: strain NS80567
;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
;Accession: T42574
;Gen. Virol. 79, 1197-1203, 1998
;Telord, E.A.; Watson, M.S.; Perry, J.; Cullinane, A.A.; Davison, A.J.
;Title: The DNA sequence of equine herpesvirus-4.
;Reference number: Z22173; MUID:98264497; PMID:9603335
;Accession: T42574
;Status: preliminary; translated from GB/EMBL/DBJ
;Molecule type: DNA
;Residues: 1-1208 <TEL>
;Cross-references: EMBL:AF030027; NID:92605950; PIDN:AAC59547.1; PID:g2605975
;Experimental source: strain NS80567
;Genetics:
;Gene: 31
;Superfamily: herpesvirus DNA-binding protein
;Keywords: DNA binding

Query Match 56.9%; Score 3583; DB 2; Length 1208;
Best Local Similarity 56.4%; Pred. No. 7.4e-264;
Matches 689; Conservative 194; Mismatches 305; Indels 34; Gaps 13;

1 MENTQKTVTVPTPLGPGVYACRVEDDLLEISFLAARSTDSLALLPLMNLTVKTFPS 60
2 MESAPKTVSLPVSPGLGVYAIQNTFMEATEALTMAARSIDSDLAVLPIRGTLVEQTFT 60
3
61 SLAVSGARTTGLAGITIKLTSTHFYSPVFPVHGGKVLPPSSAAPNLTRACNAARERF 120
61 NVAVAGSKTTLGGAGITLTPSHFTPNAPVYGVSGVFGASSKAPNLTRACELARRF 120
121 GFSRCQGPVVDGAVETTGACITLGLSEPTNTLYLVWLTALKEAVFMVNCVFLHYGLDI 180
121 GFSPFSPVDNAVETSGEICASLNSPETTLYLWVTTETKEMVYMCNTLHYGGTST 180
181 VHNHGVIRIPFPVOLFMVDNRLVPPDFTNTHRSIGEGFVYPPFTNTGLCHLHDC 240
181 VTHGOEAVKIPIYVOLVMPDNRLAAPPNSKHSIGDEFVYKPFNSDLCHLHGY 240
241 VIAPMAVALRVNVTAVAGAAHLADENHEGAVLPDDITYTYFQGS---SGTTTARGAR 298
241 VLGPAVALRVNLDVARGAAHLADENHEGSLVPQDVTFTLFDPSAQGTSGKSGRTQR 300
299 RNDVNSTKSPSPSGGFERRLASIMAADTALHAIEVFNITGIEETPTDIKEWPMFIMGST 358

```

```

Db 301 QGD-GSGLKNGSSGIERRLASIMAADTALSVDSIMGAGVYDTELPSEVDLP1-LSVGDD 358
QY 359 LPRNLALGSGYARVAGVIGAMVSPNSALVLTVEVDSCHTEAKGGPGSPFRFYFAGP 418
Db 359 RERLEALGAVASRLSLGVAMVFSANSLVTEVDDGGPADGKAS-NPSYHRFYLIAAP 417
QY 419 HLAANPOTDGDGHVLS-----SQSTGSSNTEPSVDYLALICGFGAPLILARLLFYLERCD 472
Db 418 YVAGNPOTDGDGVLGHTADQPAAPINGNOEFLDYALACGFCOLLARILFYLERCD 477
QY 473 AGAFTG-GHGDALKYVTGTFDSIPCSLCEKTRPVCAHTTVHRLRQMRPFGQATQPI 531
Db 478 ACTFGGNETDALRYLANLSEVPCGLCTPATRPACHTTLHRLRQMRPFGTTPVRBI 537
QY 532 GVTGWNVSQVSDCDPLGNVAPYLILRKPGDQTEAAKATMQDTYRATLERLFTDLEERLL 591
Db 538 GIFTGWNSTVSDCDVLGNVASYGALKRPND-NEAPKSIMQDTYRATMERLVNDLEAKLI 596
QY 592 DRGA-----PCSEGLSSVIVDHPFERRILDTLRARIBQTTTFQMKVLVETDYKIREGL 646
Db 597 DKEALAHAGTCAS--TGUVKQASFINLLSTIKDITEGAABEQFMRLLVEVDFKIREGL 654
QY 647 SEATHSMALTFDPYSGAFCPITNFLVKRTHLAVVODLALSOCHCVFGQOEGRFRNQF 706
Db 655 ADANTHMSISLDPYSSFCVTSPLSRRTIFAVLQDLVLSQCHCLFYGOSVEGRFRNQF 714
QY 707 QPVLRERRFVDFLNGGPISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGLARVSVRI 766
Db 715 QPVLRERRFVDFLNGGPISTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGLARVSVRI 766
QY 767 RDIRVKNRVVFGSCNTNLSEAAARLVGLASAYORQEKRVDMHGLGALFLKQPHGLLFP 826
Db 775 RDLIRKNRVLFSGGANNSEAAARVAVAGASAYRPERKGSNLLNGAVGLVQKPHGLLFP 834
QY 827 RGNPNSKSPNQWFTWTLQORNPADKLTHEETITIAVKRFTTEEYAAINFINLPPTCI 886
Db 835 RGHPPGIDTPNPQWFTWTLQORNPADKLTHEETITIAVKRFTTEEYAAINFINLPPTCI 886
QY 887 GELAQFYMANLILKYCDHSOYLINTLTSITGARRPRDPSSVLHWIRKDVTSAADETQA 946
Db 895 GELAQFYANLVLYKCDHSOYFINGLTAVWGSRPRDPAVLAWTDRIINGAADVEPAA 954
QY 947 KALLEXTENLPELWTTAFTSTHVRAMQORPMVLGISIKYHGAAGNNRVFOAGNWSG 1006
Db 955 QEVQLQGLGNPAATGTFSTNNRVYMPQRPVWVIGLSIKYNGSAGNNRVFOAGNWSG 1014
QY 1007 LGGKQNVCLPFTDRTRRTIACPRGGFICPVTPGSSGNRETTLSDOVRGIIVSGAMVQ 1066
Db 1015 LGGKQNVCLPMAFDRTRRVFLACPRVGFCEAGGFTGVRENTLSEQVRGIVSEGGPMVQ 1074
QY 1067 LAIYATVAVGARAQAHAFDOWLSLTDDEFLARDELDHDIQTLETPTWVEGALEAV 1126
Db 1075 TAVFSVLTLGARTOHLAVDDMIGLVDDDEFLAASLDALNATVWQF-GEWSVEAAQDMI 1133
QY 1127 KILDEKTTAG---DGETPTNLAFFDSC---EPSHDTTNSVLNISGNSISTVPGKLRP 1181
Db 1134 RTMDAQTNMGVSVTGDG-----AFDFGACVGDANOSSTTFNMGPA--SSAPAGQKRP 1185
QY 1182 EDELFDLSGIPKIKHGNITMEM 1203
Db 1186 PDDILFDMGAPPEKKSGLTFTDM 1207

RESULT 4
DNBKS
DNA-binding protein - human herpesvirus 1 (strain KOSI.1)
;Species: human herpesvirus 1
;Note: host Homo sapiens (man)
;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 16-Jul-1999
;Accession: A28601
;Gao, M.; Bouchey, J.; Curtin, K.; Kriple, D.M.
;Virology 163, 319-329, 1988
;Title: Genetic identification of a portion of the herpes simplex virus ICP8 protein re

```

176 GGLDIIVHNGDVIRIPLFPVQLFMPDVNRVLPDPENTHRSIGRGVYPTFPYNTGLCH 235
 179 GGSDDKVTIGGAHVIRIPVYQLFMPDPSRVIAEPNANHRISIGKFTYPLFPFNRPLNR 238
 236 LTHDCVIAPMAVALRVNVTAVAGAAHAFDENHEGAVLPDPITYTYFQSSSSSTTTAR 295
 239 LFEAVVGAVALRCRNVDAVARAAAHAFDENHEGAPADPITFTAFASQG--KTPR 296
 296 GARRNDVNSTKPSGSGGPERLASIMAAADTALHAENVFNIGIYETDTEKEMPMFGM 355
 297 GGR-----DGGKGAGGFEQRLASVWAGDAALALESVMAVDFPPTDISAMPLFEGQ 351
 356 EGTLPRLNALGYSYARVAGVIGAMVFPNSALYLTVEEDSGMTAKDGGPSPFNRPYQF 415
 352 DTAARANAVALGAYLARAAGLVGAMVFNSTGALHLTEVDVADGAPADPKHSHK-PSFYRFLV 410
 416 AGPHLAANPQDRDGHVL-----SSOSTGSSNTPERSVDVILALICGFCAPLLARLLFYL 468
 411 PGTHVAANPQDRGHHVFGPEGPPTAPLVGGTQ--EPAGEHLAWLCCGFPALLAKMLFYL 469
 469 ERCDAGFTGGHG--DALKYVTGTFTDSEIPCSCIKHTRPVCAHTVHRLRORMPRFGOAT 527
 470 ERCDGAVIVGQEMDVFRYVADSNQDTPCNLCFTDTHACVHTTLMRLRAHHPKFSAA 529
 528 ROPIGVGTWMSQYSDCDPLGNAPYLLIRPGQOTEAAKATMODTYRATLERLFDLEQ 587
 530 RGALGVGTWMSMYSDCDVLGNAYAFSAALKR--ADGSETARTIMOETRYAATERVMALET 588
 588 BRLLDRGAPCSSEGLSSVIVDHTFRILDLTARIEQTTTQPMKVLVETRDYKIREGLS 647
 589 LOYVDQAVPTAMGRLEITIINREALHTVNVNVOVDREVEQLMRNLVEGRKFKPDGLG 648
 648 EATHSMALTDPYSGAFCPITNFIKTHLAVQDLALSOCHCVFYGOQVEGRNFRNQFQ 707
 649 EANHMSLTLDPYACGFCPLLQLLGRSNLAVQDLALSOCHGVFAGQSVGEGRNFRNQF 708
 708 PVLRRRVDLFGNGFISTRTITVTLSEG--PVSAFNPTLGDAPAGRTFDGLARVSEVI 766
 709 PVLRRRVMDVFNNGFLSAKTLTVALSEGAAICAPSLTAGOTAPAESFEGDVARVTLGPP 768
 767 RDIRVQRVWFGNCTNLSEAAARLVGLASAYORQEKRVDMHLHGALGFLKOPHGLRPP 826
 769 KELRVKSRVLFAGASANASBAKARVASLOSAYQKPKRVDDILLGPIGLFKQFHAIFP 828
 827 RGMPPNSKSNPQWFTLLQONQPADKLTHEITTTAAVKRFTTEYAAINFILPPTCI 886
 829 NGKPPGNSQNPQWFTALQONQLPARLLSREDIETIAFKTFLDYGAINFILAPNNV 888
 887 GELAQFYMANLILKYCDHSOVLINTLTSIITGARPRDPSSVLHWIRKDVTSADLETQA 946
 889 SELAWYTMANQLLYCDHSYFINLITAI TAGSRPPSVQAAAAW---SAQGGAGLEAGA 945
 947 KALLEKTENIPELMTTAFSTHLVRAAMNORPMVYLGISISKYHGAAGNNRVFQAGNWAG 1006
 946 RALMDAVDAHFAGAWTSMFASCNLLRPVMAARPMVYLGISISKYHGAAGNNRVFQAGNWAS 1005
 1007 LMGKKNVCLFTDTRTERRFIACPGGFCIPVTPGSGNRETTLSQVQRGIIVSGGAMVQ 1066
 1006 LMGKKNACPLIIDRTKFKVLAICPRAGFVCAASSLOGGAHESLCEQLRGIIISGGAAVA 1065
 1067 LAIYATVVRVAGARAHMADDDMLSLTDDFLARDLEELHDQIIQLETPTWVEGAL--- 1123
 1066 SSVFVATVKSILGPRTOQLQIEDLALLEDEYLSSEMMELTARALENGENSTDALEVA 1125
 1124 -EAVKIIDEKTAGDGETPNNLAFND--SCPSHDTTGNVLNIGSNISGTVGPKLKP 1180
 1126 HEABALVQSGNAGE-----VFNFGDFOCE-----DDNATPFGPGAPGAPAGRKEA 1173
 1181 PEDELFDLSGIPIKHGNITMEM 1203
 1174 FHGDDPEG-EGPDPKKGDLTDM 1195

RESULT 6

DNBEHF
 DNA-binding protein - human herpesvirus 1 (strain F)

C;Species: human herpesvirus 1

A;Note: host Homo sapiens (man)

C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 07-Jun-1996

C;Accession: D29242

R;Hammerschmidt, W.; Conraths, F.; Mankertz, J.; Pauli, G.; Ludwig, H.; Buhk, H.J.

Virolgy 165, 388-405, 1988

A;Title: Conservation of a gene cluster including glycoprotein B in bovine herpesvirus 1

A;Reference number: A94381, MUID:88306231, PMID:2841793

A;Accession: D29242

A;Molecule type: DNA

A;Residues: 1-1196 <HAM>

A;Cross-references: GB:M21629

C;Superfamily: herpesvirus DNA-binding protein

C;Keywords: DNA binding

Query Match 48.4%; Score 3047; DB 1; Length 1196;

Best Local Similarity 49.8%; Pred. No. 4.7e-223;

Matches 609; Conservative 199; Mismatches 367; Indels 48; Gaps 17;

QY	1	MENTQKTVT---VPTGPLGYVY--ACRVEDLDLERISFLAARSTDSDLALLPLMRNLVE	55
DB	1	METKPKTATTIKVPPGLGYVYARACPSGIEL--LALLSARSGSDVAVAPLVGLTVE	58
QY	56	KTFSTSLAVVSGARTGLAGGITLKITTSFYPSVVFHGGKHVLPSSAANPLTRACNA	115
DB	59	SGFEANVAVVGSRTTGLGTAVSLKLPESHYSSVYVFHGRHLDPSTQAPNLTLCER	118
QY	116	ABERFGFCRCQPPVDGAVETTGAEICTRLGLEPENTLYLVVLTALFKEAVFMCNVELHY	175
DB	119	ARRHFGSDYTPRGDLKXHTTGALCERLGLDDPRALLYLVVTGFKFAVSINNTFLHL	178
QY	176	GGLDIVHNGDVIRIPLFPVQLFMPDVNRVLPDPENTHRSIGRGVYPTFPYNTGLCH	235
DB	179	GGSDDKVTIGGAHVIRIPVYQLFMPDPSRVIAEPNANHRISIGKFTYPLFPFNRPLNR	238
QY	236	LTHDCVIAPMAVALRVNVTAVAGAAHAFDENHEGAVLPDPITYTYFQSSSSSTTTAR	295
DB	239	LFEAVVGAVALRCRNVDAVARAAAHAFDENHEGAPADPITFTAFASQG--KTPR	296
QY	296	GARRNDVNSTKPSGSGFERRLASIMAAADTALHAENVFNIGIYETPTDKEWPMFIGM	355
DB	297	GGR-----DGGKGAGGFEQRLASVWAGDAALALESVMAVDFPPTDISAMPLFEGQ	351
QY	356	EGTLPRLNALGYSYARVAGVIGAMVFPNSALYLTVEEDSGMTAKDGGPSPFNRPYQF	415
DB	352	DTAARANAVALGAYLARAAGLVGAMVFNSTGALHLTEVDVADGAPADPKHSHK-PSFYRFLV	410
QY	416	AGPHLAANPQDRDGHVL-----SSOSTGSSNTPERSVDVILALICGFCAPLLARLLFYL	468
DB	411	PGTHVAANPQDRGHHVFGPEGPPTAPLVGGTQ--EPAGEHLAWLCCGFPALLAKMLFYL	469
QY	469	ERCDAGFTGGHG--DALKYVTGTFTDSEIPCSCIKHTRPVCAHTVHRLRORMPRFGOAT	527
DB	470	ERCDGAVIVGQEMDVFRYVADSNQDTPCNLCFTDTHACVHTTLMRLRAHHPKFSAA	529
QY	528	ROPIGVGTWMSQYSDCDPLGNAPYLLIRKPGQOTEAAKATMODTYRATLERLFDLEQ	587
DB	530	RGALGVGTWMSMYSDCDVLGNAYAFSAALKR--ADGSETARTIMOETRYAATERVMALET	588
QY	588	BRLLDRGAPCSSEGLSSVIVDHTFRILDLTARIEQTTTQPMKVLVETRDYKIREGLS	647
DB	589	LOYVDQAVPTAMGRLEITIINREALHTVNVNVOVDREVEQLMRNLVEGRKFKPDGLG	648
QY	648	EATHSMALTDPYSGAFCPITNFIKTHLAVQDLALSOCHCVFYGOQVEGRNFRNQFQ	707
DB	649	EANHMSLTLDPYACGFCPLLQLLGRSNLAVQDLALSOCHGVFAGQSVGEGRNFRNQF	708
QY	708	PVLRRRVDLFGNGFISTRTITVTLSEG--PVSAFNPTLGDAPAGRTFDGLARVSEVI	766
DB	709	PVLRRRVMDVFNNGFLSAKTLTVALSEGAAICAPSLTAGOTAPAESFEGDVARVTLGPP	768

DNA-binding protein - bovine herpesvirus 2 (strain BMV)
 C;Species: bovine herpesvirus 2
 C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 07-Jun-1996
 C;Accession: A29242
 R;Hammerschmidt, W.; Contraths, F.; Mankertz, J.; Pauli, G.; Ludwig, H.; Buhk, H.J.
 Virology 165, 388-405, 1988
 A;Title: Conservation of a gene cluster including glycoprotein B in bovine herpesvirus 2
 A;Reference number: A94381; MUID:88306231; PMID:2841793
 A;Accession: A29242
 A;Molecule type: DNA
 A;Residues: 1-1186 <HAM>
 A;Cross-references: GB:M21628
 S;Superfamily: herpesvirus DNA-binding protein
 S;Keywords: DNA binding

Query Match 46.6%; Score 2933; DB 1; Length 1186;
 Best Local Similarity 48.7%; Pred. No. 2.2e-214;
 Matches 585; Conservative 188; Mismatches 373; Indels 56; Gaps 11;

2Y 1 MENTOKTIVT---VPTGPIGVYACRVEDLDBEISFLAARSTDSLALLPLMRNLTVEKT 57
 Db 1 MENKQATATVVKSPGPGVYVARRLPPEGLTELALLSARSADSDTAVLPIAGLTVESG 60

2Y 58 FTSLAVSGARTTLAGAGITLKLTTSHFYPSVFVHGKHVLPSSAAPNLTRACNAAR 117
 Db 61 FDVNAVVGSRITGVGTGVSCLKMPSHYASAVVFGGRHLAPSSAAPNLSLDCRAR 120

Y 118 EFGFSRCOGPPVDGAVETTGAEICTRLGLEPENTILYLWLTALFKEAVFMCNVLHYGG 177
 b 121 VQGFSSFKPKCEAGEGETTGALCEHLGNLNEISLLYVIAEGKPAVYISNLTLMGG 180

Y 178 LDIVINHGDIVIRLPLFPVQLEMPDVNRLVPPNTHRSIGEGFVYPTPTNGLCHLI 237
 b 181 VCTVTIAGEEVRIRIPIQLQMPDYCRVADPFNDRHRAIGEPAYPLFPFNKASLL 240

Y 238 HDVCIAPNAVALVRNVTAVARGAAHLAFDENHEGAVLPDITTYFQSSSGTITARGA 297
 b 241 FGAAGVPAALVARNVDVAPAAAHAFDENHEGAALPAITTAEDPTQS-----XAG 295

Y 298 RENDVNSTSKSPSGFRRILASIMAAATLHAENVFNTGISTPTDIKWPFIWEG 357
 b 296 HRNPREC-----GGFFORLASVWAGDAALALEINSMAVFEPPTDITGWMLTQES 349

Y 358 TLPRNLALGYSYARVAGVIGAVFSPNALYLTEVEDSGMTEAKDGGGSPFNFYQFAG 417
 b 350 TAARAASIGAVLGRAAGLVGNVSSNALHLTEVDVADGAPDKD-PTKPSFYFFLVP 408

Y 418 PHLAANPOTDRDGHVLSQS-----TGSSNTEPSVDYLALICGFGAPLLARLLFYLERC 471
 b 409 TYVAANPOLDRDGRVWAGHEGRPIVPIVGGNHEFTCEHLATLQGFPELLAKMLYLERC 468

Y 472 DAGAFTG-GHGDAKYVTGTDSSIPCSLCEKHTPVCATHTVHLRQMRPFGQATROP 530
 b 469 DGGVILGRPEMDTFKYVSDSAHTDVPCCLSLDRHSCAHTTLLRLRARHPKFTSTTGA 528

Y 531 IGVFGTMNSQYSCDPLGNAPYLILRKPGQTEBAKATMDTVRATLERLFDLEQERL 590
 b 529 IGIFGWMSAYSDDVLGNAYASFSAIKEM-DVQETARAIQGETVRSVERVMAELENLY 587

Y 591 LDRGAPCSSEGLSSVVDHPFRILDLRIARISQTTQTFPMKVLVETRDYKIREGLSEAT 650
 b 588 IDAAVPTSPAKLEIITGREALQTVSNVKVQVDGVEAQLMRALVEGRGFRFEALGEAN 647

Y 651 HSMALTDPYSGAFCPITNFIKVKTHLAVVDLALSOCHVCFVQGVQVEGRNFRNQPV 710
 b 648 HAMSILDPHASVCPQLQLQMLGRSNLAVQDLALSOCHGVFEGQAVEGRNFRNSQPV 707

Y 711 RRRFVDLPNGFISTRTITVLTSEGP-VSAPNPTLGQDAPAGRTFDGDLARVSEVIRDI 769
 b 708 RRRVLDNFNGFLSARTLTVALTDGACISAPGLVSGQHAASGSEFGEDVARVNLGPFKEI 767

Y 770 RVKRVVPSGNTNLSSAARLVGLASVQREKRVDMHLGALGHLLKQFHGLLPFRGM 829

Db 768 RVKSRVLFAGAGPAASEAARARIAGLSAQVQSKDKRVDILLGLPGLMFKQFHATLPNGK 827

QY 830 PPNKSPNPQFWTLQORNQMPADKLTHEEITIAAVKRETEEYAAINFNLPTTCIGEL 889

Db 828 PPGSDNPNQFWTLQORNLPARLLSREDISLIAVFKGFSVEYGAGNFWNLPNNISEL 887

QY 890 AQFYMANLILKYCDHSOYLNTLTSITGARRPRDSSVLHWIRKDVTSAADTETQAKAL 949

Db 888 AMYMANQILKYCDHSYTFINTUTALTAGRRRPPNQAASAAAPRGGT---ELEAQRSV 944

QY 950 LEKTENLPETLTAFTSTHLVRAAMNORPMVLGIGISIKYHGAAGNNRVFOAGNWSGLNG 1009

Db 945 VANPDGHPGAWTTFEASCNLLRPWMATRPNVVLGLSISKYGVAGNDRVFOAGNLANILG 1004

QY 1010 GKNCVPLFTFDRTRRRIIACPRGPFICPTVGPSSGNRETTLSQVGRGIIVSGAMVQLAI 1069

Db 1005 GKNACPLLIIFDRTRKVIACPRAGFVCAAVSAGSAGHESLCEQLAAIIAEGGATVASDV 1064

QY 1070 YATVVRAGARAHMAFDWLSLTDDEFLARDLEELHDOIITLETPTWTEGALEAVKIL 1129

Db 1065 FAAAASLGRVQLOQIEDWLALLEDEYISEEMWELAGRALERGGGSESLDAALDVAREA 1124

QY 1130 DEKTT-----AGDGETPTNLAPNFD-----CEPSSHDTTSEN 1160

Db 1125 EAMVTRHVDAAETTFDGAFAEDCPADAGLAVHLSQRRRLACSDILFGDAPAEKNDLTLD 1184

QY 1161 VL 1162

Db 1185 ML 1186

RESULT 9
 JQ0846
 DNA-binding protein - equine herpesvirus 1 (fragment)
 C;Species: equine herpesvirus 1
 C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 23-Feb-1997
 C;Accession: JQ0846
 R;Bell, C.W.; Whalley, J.M.
 submitted to JIPID, January 1991
 A;Reference number: JQ0846
 A;Accession: JQ0846
 A;Molecule type: DNA
 A;Residues: 1-375 <BEL>
 C;Superfamily: herpesvirus DNA-binding protein
 C;Keywords: DNA binding; nucleus

Query Match 17.4%; Score 1094.5; DB 2; Length 375;
 Best Local Similarity 55.5%; Pred. No. 3.1e-75;
 Matches 212; Conservative 57; Mismatches 98; Indels 15; Gaps 5;

QY 828 GMPNPKSPNPQFWTLQORNQMPADKLTHEEITIAAVKRETEEYAAINFNLPTTCIG 887

Db 2 GHPPGIDTPNPQFWTLQORNQMPARLLSKEDIETITAIKFSDEYSAINFINLTPNIG 61

QY 888 ELAQFWMANLILKYCDHSOYLNTLTSITGARRPRDSSVLHWIRKDVTSAADTETQAK 947

Db 62 ELAQFYFANLVKILKYCDHSOYFINGLTAVVGRSFRDPAVLAVIDETTINGAADVEPAQ 121

QY 948 ALLEKTENLPETLTAFTSTHLVRAAMNORPMVLGIGISIKYHGAAGNNRVFOAGNWSGL 1007

Db 122 EYLQRLGSPNPAWGTFTSTNMVYVMDQRPVVGISISKYSGAGNNRVFOAGNWSGL 181

QY 1008 NGGKNCVPLFTFDRTRRRIIACPRGPFICPTVGPSSGNRETTLSQVGRGIIVSGAMVQL 1067

Db 182 NGGKNCVPLMAFDTRFVLACPRAGFVCAAVSAGSAGHESLCEQLAAIIAEGGATVASDV 241

QY 1068 AITYAVRVAAGARAHMAFDWLSLTDDEFLARDLEELHDOIITLETPTWTEGALEAVK 1127

Db 242 AVFAAVLHALGARQHLAVDDWIGLVDDDEFIASSLDALNATVVDQF-GEWSVRAAQELVK 300

QY 1128 ILDEKTTAG---DGETPTNLAPNFDSC---EPSSHDTTSENVLNISGNSISGTVPLGRPP 1181

Db 301 NMEAQTTAGAAAGEG---AFDFGACVGTDPQGSTA---FNGGLAMAAAPAGQKSL 352

QY 1182 EDELFDLGSGPIKHGNTM 1203
 Db 353 PDDILFDGAPPEKSGLTFFDM 374

RESULT 10
 QOE47
 DNA-binding protein - human herpesvirus 4 (strain B95-8)
 C:Species: human herpesvirus 4, Epstein-Barr virus
 C:Date: 25-Feb-1995 #sequence_revision 25-Feb-1995 #text_change 16-Jul-1999
 C:Accession: A3045; A03791; S33057
 R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
 Mol. Biol. Med. 1, 21-45, 1983
 A:Title: Sequence analysis of the 17.166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
 A:Reference number: A93065; MUID:85035713; PMID:6028285
 A:Accession: A43045
 A:Molecule type: DNA
 A:Residues: 1-1128 <BAN>
 A:Cross-references: EMBL:V01555; NID:G59074; PIDN:CAA24808.1; PID:91334916
 R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H
 Nature 310, 207-211, 1984
 A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
 A:Reference number: A03794; MUID:84270667; PMID:6087149
 A:Contents: annotation; protein coding region
 C:Superfamily: herpesvirus DNA-binding protein
 C:Keywords: DNA binding

Query Match 10.6%; Score 669; DB 1; Length 1128;
 Best Local Similarity 23.2%; Pred. No. 4.8e-42;
 Matches 292; Conservative 186; Mismatches 537; Indels 246; Gaps 45;

2Y 13 GPLGYVACVEDELDEEISFLAARSTDSALLPLMRNLTKETPTSSLVAVSGARTG 72
 Db 18 GPCGYIYFPLATYPLREVATLCTGVAGHRCITVPLLCITVEPGFSIN---VKALHRRP 74
 2Y 73 LAGAGITLKTTHSHFYPSVVFHGGKHVLPSSNAPNLTACNAARRPGFS-----RCQ 126
 Db 75 DPNCGL---LRATSHRDYVFINAHMVPPIFEGPGLEALCGTREVFGDYAYSALPRES 131
 2Y 127 GPVDAVETGAEICTRLGLBEPNTILVLTALFKEAVFNCVNFVHVGGLDIHINHG 186
 Db 132 SKPGDFPE-----GLDPSAVLGAVATEAFKERLYSGNLVAPSLKQEVAVGQS 181
 2Y 187 DVIRIPFPVQVMPVNVRLVDDPNTHRSIGEGVPYTPFYNTGLCHLIHDCVIAPMA 246
 Db 182 ASVRVPLDYKEVPFEGVPQ-----RQFNSDLKRCWHEALYTGLA 222
 2Y 247 VALRVNVTAVARGAHLAFDENHGAVALPPDITYTFQSSSSGTTTARGARRNDVNSTS 306
 Db 223 QALRVARVCKL-----VELLEKQSLQDQAKVAKVAPLKEFPAST 261
 2Y 307 KSPSGGPFERRLASIMADT-ALHAEIFNTGIYE---ETPTDIK--EWPMPFIMEGTLP 360
 Db 262 IGHPPDSG-----ALMIVDSAACELAVSYAPAMLEASHETPASLVYDSWPLFADCEGPEA 315
 2Y 361 RNLALGSYTARVAGVIGAVFSPNGLALYTEVDSGMTKADGGCPSPFNFPYAG---417
 Db 316 RVAAHLRYNVAAPHVSTQIFATNSVLVY-----SGVSKSGQGKESLFSFYTHGLGT 370
 2Y 418 -----PHLAANPQTDRGDHVLSSQSSNTSFSDYDALICGFCAPILARLLF 466
 Db 371 LQEGTWDPCCRRPCFSGWGPDTG-----INGPQN--YAVEHLVYAASFNLLARYAY 422
 2Y 467 YLERCDAGFTGGHGDAL-----KYTGTFDSEIPSLCEKHTRPVCAHTTVHRLQR 519
 Db 423 VLQPCQ-----GQKSLTVPETGSYAGAAAFPM-CSLCEGAPAVCLNTLFFRLDR 475
 2Y 520 MPRFQATRQFIGVFGTWNQSDCDPLNVPAYLIL---RKPGDQTEAAKATWQTYR 575
 Db 476 FPPVMSVTRQRPYVTSAGSGSYNETDFLGNFLNFDKEDDQCRDDEPRYTWQNLNLL 535
 2Y 576 ATLERLFDLQERLLDRGAPCSSEGLSSVIVDHPFTFRILDTLRARIEQTTQFMKVLV 635

Db 536 ERLSRLGIDAECKLEKPHGP-----RDFVKMFQDYDAADAEVQFMASMA 582
 QY 636 ETRDYKIREGLSEATHSNALTDPYSGAFCCPITNFWLVRTHLAVQDLALUSQCHCVYQG 695
 Db 583 K-NNTYKDLVKRSCYHVMOYSCNPPAFQACPIFTQLFYRSLTLTLOISLPICMC-----636
 QY 696 QVEGRNFRNQFVLRRRFDLFGNGF-----ISVRSITV-----TLSEGPVSAFNPITLG- 745
 Db 637 -----YEND-NPGLQGSPPEMLKGHYOTLCTNFRSLAIDKGVLTAKKVVHGEPTCDL 689
 QY 746 -----QDAPAGRTFDGLARVSEVIRDIRKVRVVFSGNCTNLSEAAARLVGLASA 798
 Db 690 PLDAALOGRVYGRRLPVRMSKVLMLCPRIKIKRVRVFTGNAALQNS-----FIKS 742
 QY 799 YQREKRVDMHLHAGLGFLLKQFHLFPFGMPNPKSPNQWFLLQRMQPADK-LTH 857
 Db 743 TTRRENYI--INGPYMKFLNTYHKLTF-----PDKLSSLYLWHNFRSRRSVVPVSGASA 795
 QY 858 EBITTIAA-VKRFTEEVAANFINLPTTCIGELAQFYMANLILKYCDHSQYLINTLTSII 916
 Db 796 EYSDLALFVDGGSGRAHEESNVIDVPGNLVTAKORLNNAILKACQQTQFYISLIQGLV 855
 QY 917 --TGARRPRDPSSVLRKDVTSAAADIETQAKALLEKTENLPELWTTATFTSTHLVRAAM 974
 Db 856 PRTQSVFARDYPHVLG--TRAVESAA-----AYAEATSSLTAT-TVVCAATDCLSQVC 905
 QY 975 NORPMVVLGISISKYHGAAGNRRVFOAGNWSGLNGKNVCPLFTFDRTRRFTIACPRGG- 1033
 Db 906 KARDVWTLFVTINKYTGWNGNNQIFQAGN-LGYFMGRGV-----DRN---LLOAPGAGL 955
 QY 1034 -----FICPTVGPSSGNR-ETILSDQVRGIVSGAMV-----QLAIYATV 1073
 Db 956 RKQAGSGMRKKVFATPTGLTVKERTQAATTYEIENTRAGLEAIIISQKQEDCVFVV 1015
 QY 1074 VRVAGARQAHFADMLSLT--DDEFLARDLELHQIITQLET-----PMTVEGALEA 1125
 Db 1016 CNLVDMAGEACA-----SLTRDDAEVLLGFRFVSLADSVLETATASSGIETAEARDP 1070
 QY 1126 VKILDEKTTAGDGETTNLAFNDFSCPSHDHTTSNVLNTSGSNISG-----STVPGL 1177
 Db 1071 L-----EGWVGPGGAQDNFTSVAEPYSTASQASAGLLGGGGGSGGRKRRLATVLPGL 1126
 QY 1178 K 1178
 Db 1127 E 1127

RESULT 11
 T42922
 major single-stranded DNA binding protein - ateline herpesvirus 3 (strain 73)
 C:Species: ateline herpesvirus 3
 A:Variety: strain 73
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 19-Feb-2000
 C:Accession: T42922
 R:Albrecht, J.C.; Fleckenstein, B.
 submitted to the EMBL Data Library, August 1998
 A:Description: Primary structure of the herpesvirus ateles genome.
 A:Reference number: Z22274
 A:Accession: T42922
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1128 <ALE>
 A:Cross-references: EMBL:AF083424; PIDN:AAC95531.1
 A:Experimental source: strain 73
 C:Superfamily: herpesvirus DNA-binding protein

Query Match 10.1%; Score 633; DB 2; Length 1128;
 Best Local Similarity 22.4%; Pred. No. 2.6e-39;
 Matches 274; Conservative 194; Mismatches 514; Indels 240; Gaps 41;

QY 11 PTGPGYVYACRVEDLDLEISFLAARSTDSALLPLMRNLTKETPTSSLVAVSGART 70

Db 18 PVGAGGIYVYKGGPPFQEQASLLGNRNNGADAVSLPLSLGLTVEANFSFNKAVHKID 77
QY 71 TGLAGGITLKLTTSHFPYVVFHGGKHVLPSSAAPNLTRACNAARERFGFSRCQGPV 130
Db 78 M-----TTLVSRAVSHREAIYFNTNFTPIFFPGGLDLSLSDARNLFGYTSFT-PRH 130
QY 131 DGA-----VETGAEICTRLGLEBENTILYLVVLTALKEAVFMCMVFLHYGGGLDIVHNG 186
Db 131 DLKDLIDIKOLYAPFYTK-----DSCFMSVVVTFESFERLYFGNLVPLISQGGKQVNGR 185
2Y 187 DVIRIPLFPVQLFMPDVNRLVPDPFNTHRSIGEGFYVTPFPVNTGLCHLHDCVIAPNA 246
Db 186 EAVKIPLYDEDLFSKSEHL-----PRFIPSVSKYLHDSLFTSTA 226
2Y 247 VALRVNNTAVAGAHLAPDENHEGAVLPDPITYYFQSSSGSTTTARGARNVNST 306
Db 227 QALRIRVESVIRAEKQSTHDOYKLAV-----VNSK- 259
2Y 307 KPSPSGGFFRRIASIMAAATA-----LHAEIFNTGI-YEETP-----TDIKEMWMP 352
Db 260 -----EFALQAVKQDASAPFVWTDCTAAELAISYGLSFLFTQEPCCALLDYTSNPIF 311
2Y 353 IGNEGTLPLNALGYSYARVAGVIGAMVSPNSALYLTVEGSGMTEAKDGGPGSPENRF 412
Db 312 DTTETBEGRIKAIQDWMAMSVHVYTHLFTSTNSVLYLKINKQTQTNKSDQV---IYNSY 368
Y 413 YQFAGPHLANPQTDGDHVLSSQSGSNTBFSVDYLALICGFGAPILARLLFYLERCD 472
b 369 FMQGLSYAAEAATQKENGDPAPFSGAVKFNNGSYTLFHLALASSFPHLLARNCYMQPCQ 428
Y 473 AGAFTGGHGDALKYVTGTFDSIPICSLCEKHTRPVCAHTVHRLRQRPVPGQATRPQIG 532
b 429 HQKSTNSYSAQVGVIVAVSELCELCQKCPAACIHTLFRUKDRPPVLSQRDDPY 488
Y 533 VEGTWNQSYDCCDPLGNVAPYVILKPGD-----QTEAAKATMQDTYATLBERLIDLEQE 588
b 489 VVTGSGQYNDLMDLGNFATF---REKEDDMVQNTCEKYTWQLIQNVVKL----- 538
Y 589 LLDRGAPCSSEGL--SSVIVDHPFTRAILDTLRARIEQTTQPMKVLVETRDYKIREGL 646
b 539 -----ATIGITEGLGSLIITDIQSLKTFREIDNVVDNEVIRKINCLVK-NKINFRETI 592
Y 647 SEATHSMALTFPPYSGAPCPTNELVLRKTHLAVQDLALSCHQCVFYGQ----- 695
b 593 KGVHVLHYCCNVFQWQPCAMFLNFYKSVLAIQDCLPV--AMIVEQDNPSMGWMPSE 650
Y 696 -----QVEGSEFN-----QPPVLRFRFVDLFGNGFISTRSITVLTSEGPVS 738
b 651 WLMHYQITWTFXSSCIDKGVLTGASHKVVHGMFCDFINVD-----S 694
Y 739 APNPTLGDAPAGRTFGDLARVSEVIRDIRVKNRVVFGNCTNLSEARARLVGLASA 798
b 695 ALN--GQIVPV--KMQVRLAKALITVPKTIKIKRIVFSN--SSMTEATQAGFI----- 742
Y 799 YQREKRVQD--MLHGALGFLKQFHLGFLFRGMPNPSKSPNQWFW--TLQRMQMPA-DK 854
b 743 --RSSTKDSYIVTGPMYKFLNSLHKVIFP-----DAKISALYLVHWTFSQKQIPLPG 794
Y 855 LTBEIETITIA-AVKRETEEYAAINFILNPTCIGELAQFYMANLILKYCHDSYLYNTL- 912
b 795 ISKENIETELANYIEAGSMHDDMMVLDIIPITLAVAKVELANNTILRTCGQTOFYATTIQ 854
Y 913 -----TSIITGARPRDPSVHLWRKDVTSADIIETQAK---ALLEKTENIPELMTTAP 964
b 855 CLLPITLQWSSATEYH-----VLH--QOSITSVDVYLSIKKQALLIVQTLKEDIATIG- 907
Y 965 TSHLVZAAAMQBMVVLGISISKYHGAAGNNRVFOAGN---MSGUNGQNVCPLETF-ED 1020
b 908 -----KHRPIVTVPLVANKYTGINGNTQIFQCNGLGFMWGRGVDRNLIPSSGFR 957
Y 1021 RTRRFIIACRGGGFIQVTC-----PSSGNREI-----TSLDQVRGIIVSGGAM 1064
b 958 RONNSAYMRKHFIMTPIVANLKRISNLNLTPEVETIRKKNVQTIFFDKDNLNIPDNV 1017

QY 1065 VOLAIYATVTVAVGARQHWAFDWMLSLTDDEFLARDLEELHQIIQTLETPTWTEGAL 1124
Db 1018 IEL-----VKLGESCENTEDDLQFYLGDIYMSDEIWSRFQLLTDSGTSPVSE---S 1068
QY 1125 AVKILDEKTTAGDGETTNLAF 1146
Db 1069 VTKLGSRK-----QEDCNLEF 1085

RESULT 12

SS5600

single-stranded DNA binding protein 06 - equine herpesvirus 2

C/Species: equine herpesvirus 2

C/Date: 27-Oct-1995 #sequence revision 03-Nov-1995 #text change 26-Aug-1999

C/Accession: S55600

R/RefSeq: E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.

J. Mol. Biol. 249, 520-528, 1995

A/Title: The DNA sequence of equine herpesvirus 2

A/Reference number: S55594; MUID:95302501; PMID:7783207

A/Accession: S55600

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-1145 <REL>

A/Cross-references: GB:U02824; NID:9695172; PIDN:AACI3793.1; PID:9695178

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995

C/Superfamily: herpesvirus DNA-binding protein

Query Match 9.8%; Score 618; DB 2; Length 1145;
Best Local Similarity 23.1%; Pred. No. 3-8e-38;
Matches 295; Conservative 182; Mismatches 540; Indels 260; Gaps 47;

QY 4 TOKIVTPTGPIGYVYACRVEDLDEETSFIAARSTDSDLLIPLMRNLTVKFTSSIA 63
Db 17 TQASV---GPGFVLYPGDTFPVEASLGNLHAGVEFSLPLSLGLTVEADPHNVK 72
QY 64 VSGARTTGLAGAGITLKLTTSHFPYVVFHGGKHVLPSSAAPNLTRACNAARERFGFS 123
Db 73-AVHKILDEP-----TVSKASAYHREVIVFANACEKPIFAGPGLGLEGLCAASQLFGYA 126
QY 124 RCGGPPVDGAVETTCG-----EICTRLGLEP--ENTILYLVVLTALKEAVMCMVFLHYG 176
Db 127 EPE-----ERAGGAARPFELADLHLLGASHAGVGVVTFESFERLYRGQLVVVES 178
QY 177 GUDIYVHNGVDVIRLFPFVQLFMPDVNRLVPDPFNTHRSIGEGFYVTPFPYNTGLCHL 236
Db 179 QIOSVRVGECDAFKVPYLDGELFAKSPCR-----ENLRY---FYHAGVSRV 221
QY 237 THDCVIAPMAVALRVNNTAVARGAHLAFDENHEGAVLPDPITYYFQSSSGSTTTARG 296
Db 222 LFEAHYTSLAQALRVDPVPGILGALERQSF---HDQVKLP-----KVVEKREFPATGHRG 273
QY 297 ARNDVNSTSKPSPGGFPERELASIMAADTALHAFLVFNLTGIYEETP-TDIKEMPFIQM 355
Db 274 A-----GDCSLTIVDSVATELAVSGLSFLFVPEQGTALLSYDKWPIPEGC 319
QY 356 EGTIPLRLNALGSYARVAGVIGAMVFPNSALYLTVEGSGMTEAKDGGPGSPFNREYQF 415
Db 320 ETPCEORVETLQFNAQAVHVSQLFSGNSVLYLARVOKQASN--RGGGENVYNSFPMG 377
QY 416 AGPHLANPQTDROG-----HVLSSQSTGSSNTEFSVDYLALICGFGAPILARLLFYLER 470
Db 378 HGLACLAEPQKENGFLSPFGVGPASALSGSN---YSLHLAYAASFSPQMLARHCYVLOF 434
QY 471 CDAGFTGGHGDALKYVTGTFDSIPICSLCEKHTRPVCAHTVHRLRQRPV-FGQATRC 529
Db 435 AQHKQSSNNSSVNYPTVTVGTAANTPCELRCGSCPASCYNTLFRILDRFPVVASVRD 494
QY 530 PIGVFGTWNQSYDCCDPLGNVAPYVILIRKFGDQTEAAKATMQDTYATLBERLIDLEQER 589
Db 495 FYVVTGVAGA-YNLDLDIAGNFANY---RDXDESSNQSEEREKEFTYMQVTQVL-----ER 545
QY 590 LLDRGAPCSSEGLSSVVDHPTFRRILDLTRARIEQTTQFMKVLVETRDYKIREGLSEA 649

```

Db      546 LSEAGICEGEGDVDAIHNIGSFLKVFKEIDGIVGEVARFINSVK-NVNVYRESIKSI 604
QY      650 TTSMALTEDPYSGACPIINFLVKEITHLAVVODLALSOCHCVFYGO----- 695
Db      605 HHVQVVCNVYQPCPFNLNLYRCVLAIVODICLPT--CMYEQENPAVGVSPGEWLK 662
QY      696 ---QVEGRNFRN-----QPQVLRFRFVDLEN-----GGFISTRSITVTLSSEG 735
Db      663 MHYQTLWTFKNSCIDKGLVLTCTEYKVVHKQFCDFDVSARGEFSCKT-QVRSRA 721
QY      736 PVSAPNPLTGODAPAGRTFDGLARVSVEVIRDIRKVRVFSNGCNLSAARLVL 795
Db      722 LMWP-----RVMKIKRUIFSPCTES-----I 746
QY      796 ASAYQORERVD--MLHGALGFLKQFGLLFPFGMPNPSKPNQFW-TLLQRMQMPA 852
Db      747 QNAFVRGTFKGDSCVSGPYMRFSLTHSOLFP-----GAKISFLFHWTFSKROLFP 800
QY      853 -DKLTHEITIA-AVKRTEEYAAINFILPPTCIGELAQFYMANILIKYCDHSQY--- 907
Db      801 FPNVPRESVTELANYEQNSRLHGETSIIDVVPENFVYAKVRLNNALFRACGQTQFYAT 860
QY      908 LINTLTSIITGARRPROPSSVYLHWRKDV-SAADIEQAQKALLEKTENLPELWTTAFTST 967
Db      861 TIHCLTKIQTVPABEYPHAL-----GARGVADVGYLGAARELT--VP---TVQCTSR 909
QY      968 HLVRANMQRMMVLGISIKYHGAAGNNRVFQAGN---MSGLMGGKVVCPLETFDTRTR 1024
Db      910 DNICEVGKCRPIVTLPLVKNKYTGVTGNSQIFOCANLGYFGRVDKQLIP----- 960
QY      1025 FIIACPRGGICPTVGTSSGNRE-----TTLSDOVRGIIIVGGAMVOLAIYATVRAVGA 1079
Db      961 ----DAGSF--KKQGVSTMRKHVFWMTPLSDHLLARSVQGAUV-----APEISGVR 1007
QY      1080 RAQ-----HNAFDDWLSLT-----DDEF-----LARDLEELHDOII 1110
Db      1008 RVQOILSDGDNPHVIRDVWLQVLSGSECRSVSEYDLEVMQYVIFAGDVSE-RLQL 1066
QY      1111 QTLTPTWVSEALEAVKILDEKTAGDGETPTNLAFNFSCE-----PSHDTTSNVLN 1163
Db      1067 SDLGDMSEWALSVL-----GEEEDPLGGELEFEKVEDAECLGHPQODEFALAPQ 1117
QY      1164 IGSNTISG-STVPGLKR 1179
Db      1118 AAAPQVSGSSVAGKKR 1134

RESULT 13
DNBEM1
DNA-binding protein - saimiriine herpesvirus 1 (strain 11)
;Species: saimiriine herpesvirus 1
;Note: host Saimiri sciureus (common squirrel monkey)
;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
;Accession: G36806
;Albrecht, J.
submitted to the EMBL Data Library, January 1992
;Description: Primary structure of the herpesvirus saimiri genome.
;Reference number: A36806
;Accession: G36806
;Molecule type: DNA
;Residues: 1-1128 <AB>
;Cross-references: GB:X64346; NID:960320; PIDN:CAA45629.1; PID:G60327
;Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.; Newman, C.; Wi
J. Virol. 66, 5047-5058, 1992
;Title: Primary structure of the herpesvirus saimiri genome.
;Reference number: A37309; QUID:92333688; PMID:1321287
;Contents: annotation; possible protein-coding frames
;Note: neither amino acid nor nucleotide sequence is given
;Genetics:
;Gene: 6
;Superfamily: herpesvirus DNA-binding protein
;Keywords: DNA binding

```

```

Query Match      9.6%; Score 604; DB 1; Length 1128;
Best Local Similarity 22.9%; Pred. No. 4 3e-37;
Matches 275; Conservative 189; Mismatches 509; Indels 226; Gaps 40;

QY      11 PTCPLGVYACRVEOLDLIEISFLAARSTDSDALLPLMRNLTVKTTSSLVVSGART 70
Db      18 PVEPCGYIYVYKPEGFPFKEASLLGNKGASMSLPLLSDLITVESNFSFNKAVHKID 77
QY      71 TGLAGAGITLKLITTSHFYSPVFVPHGKHVLPSSAAPNLTRACNAARERFGRSCGPPV 130
Db      78 M-----TTLVVRVAYHREAVFENTDLFPIFVGPGDLILCSDARSLEGTWF--VPR 129
QY      131 DCAVETTGAEICTRLGLEPENTILYLVVTLFKCAVEMCN-VFLHYGGLDIVHINHGDI 189
Db      130 TDLRDTVDIKOLYAPFYGEDSCFMAVVVTEGKERLYFGLNLPVPIIAOGLK VQINGREAV 188
QY      190 RIPLFPVOLFMVDVNRVLPDPENTHRSIGSGFVYPPFYNTGLCHLHCVIAPMAVAL 249
Db      189 KIPLVDEDLFSKSHL-----PRFYIPSVSKYLHDSVFTSIAQAL 229
QY      250 RVNVTAVARGAHLAFDENHEGAVLPDITTYTFQSSSGTITARGARRDNVNSTKPS 309
Db      230 KIRDVESVIRASEKSIQDOYKLAKI-----VNSK----- 259
QY      310 PSGGFERELASIMAADTA-----LHAEVIFNTG1-YEETP-----TDIKWPMFIGM 355
Db      260 -----DFSLOQVKQDASAFNVIDCIAELAMSGLSFLFLEAPQDCAVLDTYSNPITETA 314
QY      356 EGTPLRINALGSYIARVAGVIGAMVFPNSALYI TEVEDSGMTEAKDGGPGSPNFRFYQF 415
Db      315 ETEEDRIKAIQDMVAMMSVHYVTHLFSTNSVLTKINKTQTSKSEON---VNYFYMQ 371
QY      416 AGPHLAANPQDRGHVLSQSGTSSNTESVDYALI CGFAPLALRLFYLERCDAGA 475
Db      372 HGLAYADATQRENGSPAFSGAPKFGSGTTLVHLALASSFPHLLARNYVYMQFQHQK 431
QY      476 FTGGHGDALVTVTGTFDSEIPCSLCEKHTRPVCAHTVHRLRORMPR-FQATRTQPIGVF 534
Db      432 STTNANYSVQYVGTAAASDLCELCOGTCPASCILHLYFLKDRFPFVLGSRQRDPVVT 491
QY      535 GTMNSQVSDCDPIGNVAPYILIRKPGDQTBAAKATWQDTRATLERFLDLQERLLDRG 594
Db      492 G-VSGQYNDLMLGNFATF--REKEDAVQNAESEKTYWQLQNV-----EKLSTWG 542
QY      595 APCSEGLSVIVDHPPTFRRILOTLARLEQTTQFMKYLVEDRYKIEGLSEATHSA 654
Db      543 VTSGTVG-SBLITDQSFLEKTFDIDNVVSEVFKMCLVK-NNINFRITIKTVHVLH 600
QY      655 LTFDPYSGAFCPITNLFVLRKTHLAVVODLAL-----SQCHCVFYGOV 697
Db      601 YCCNVFMQAPCAMFLNLFYKSVLAIQDCLPAMTYEQDNFSGIMMPSEMLKVHY--QT 658
QY      698 EGRNFR-----NOFQVLRFRFVDLNGGFISTRITVLSBGPVSAPNPTLGO 746
Db      659 IWNFKSSCLDRGLVTSSEHKIVHTDMFCDFLN-----IDSALS-----GQ 699
QY      747 DAPAGRTFDGDLARVSVEVIRDIRVGRVVFVSGNCTNLSEAAARLVGLASAYQOEKV 806
Db      700 IVP--MKMGOVLAKALLTVPKTIKIKRIVFSN--SSMTETIQSGFIKSAT-----KDS 750
QY      807 DMLHGALGELLKQFGLLFPFGMPNPSKPNQFWTLLQRMQMPA-DKLTHEITIA- 864
Db      751 YITVGYMKFLNSLHKWMP-----PNAKISALYLWHTFSQKQLFVLPQISRENNVELAN 805
QY      865 AVKRFTTEEYAAINFILPPTCIGELAQFYMANILIKYCDHSOYLINTLSI-----ITG 918
Db      806 YVFTSSKMDHDMNVLDIIPITLLTYAKVRLNNTILRTCGQTQFYATTLQCLPTTQISA 865
QY      919 ARPR--DPS--SVLHWIR--KO-----VTSAADTETQAKALLEKTENLPELWTTAF 964
Db      866 TEYPHVLLDQINSVDHYLSIKDKHALTVQTTLKEDIATVGK----- 908

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 13:06:32 ; Search time 47 Seconds
(without alignments)
4062.725 Million cell updates/sec

Title: US-09-769-699-2

Perfect score: 6294

Sequence: 1 MENTQKTVTPGPIGVYA.....DELFDLSPGPIKGNITMEM 1203

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 15872573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A Geneseq 19Jun03.*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6294	100.0	1203	22	AA052633
2	3054	48.5	1452	23	AA025598
3	3054	48.5	1452	24	ABG73090
4	3016.5	47.9	1196	19	AA072105
5	1566.5	24.9	623	19	AA072197
6	1380.5	21.9	543	19	AA072196
7	620	9.9	1132	21	AA053126
8	183	2.9	35	20	AA026935
9	183	2.9	35	22	AA073931

10	156.5	2.5	1841	21	AA07562
11	125.5	2.0	19938	24	ABP76682
12	123.5	2.0	3210	22	ABP73958
13	123	2.0	1074	20	AA000206
14	123	2.0	1074	20	AA000188
15	123	2.0	1074	23	ABF43407
16	123	2.0	1074	23	ABF43425
17	123	2.0	1074	24	ABU13686
18	123	2.0	1074	24	ABU13704
19	122.5	1.9	901	22	ABF58224
20	122.5	1.9	2387	21	AA053655
21	121	1.9	1069	21	AA041326
22	121	1.9	3158	23	ABF62763
23	120.5	1.9	2597	21	AA053664
24	120.5	1.9	2597	23	ABG32889
25	120.5	1.9	2597	23	ABG32890
26	120.5	1.9	2597	23	ABG31316
27	120.5	1.9	2597	23	ABF47933
28	118.5	1.9	1076	22	ABG20205
29	118	1.9	1784	22	ABG24772
30	115	1.8	1150	14	AA043950
31	114.5	1.8	1496	20	AA081030
32	114.5	1.8	1496	21	AA070469
33	114.5	1.8	1496	24	ABU03498
34	114.5	1.8	1498	22	ABF11587
35	114	1.8	774	7	AA060616
36	111	1.8	4530	18	AA019629
37	111	1.8	4530	21	AA077177
38	110.5	1.8	1532	22	ABF68585
39	110.5	1.8	3070	22	AA081195
40	109.5	1.7	2630	24	ABG76186
41	109.5	1.7	7968	24	ABG76187
42	109	1.7	11096	22	AAE10129
43	108.5	1.7	1451	22	AA095868
44	108	1.7	2596	22	AA030569
45	106.5	1.7	774	16	AA086374

ALIGNMENTS

RESULT 1
AA052633
ID AA052633 standard; Protein, 1203 AA.
XX
AC AA052633;
XX
DT 18-FEB-2002 (first entry)
XX
DE Varicella-Zoster virus (VZV) ORF29p protein
XX
KW ORF29p; open reading frame 29p protein; Varicella-Zoster virus; VZV;
KW drug delivery; gene delivery; fusion protein; protein secretion.
XX
OS Human herpesvirus 3.
XX
PN WO200154709-A1.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02500.
XX
PR 25-JAN-2000; 2000US-0177901.
XX
PA (UYCO) UNIV COLUMBIA NEW YORK
XX
PI Silverstein S, Annunziato P, Gerebon A, Lungu O;
XX
DR WPI; 2001-488744/53.
XX
DR N-PSDB; ABA02175.
XX
PT Composition of matter for delivering desired compounds into eukaryotic cell, comprises Varicella-Zoster virus open reading frame 29p protein

09/769,699
2-21-04
Search Notes

either bound to an agent or operably affixed to lipid-soluble group -
 Claim 1; Fig 6; 69pp; English.

This sequence represents the Varicella-Zoster virus (VZV) open reading frame 29p (ORF29p) protein. Although ORF29p is the major DNA binding protein of VZV, the invention is based on the discovery that the VZV ORF29p protein can readily enter and exit eukaryotic cells. Accordingly, the invention relates to a composition for delivery of an agent into a eukaryotic cell, comprising the VZV ORF29p protein bound to a polynucleotide, or an organic compound. The invention also relates a composition in which the ORF29p protein is linked to a lipid soluble group that permits the protein to be anchored to a lipid membrane, and lipid vesicles containing this composition. The invention additionally relates to the use of the ORF29p protein as a facilitator of secretion of a desired protein, in which the desired protein and the ORF29p protein are recombinantly expressed as a fusion protein. The invention further encompasses nucleic acids encoding the VZV ORF29p protein and their use in detection of VZV ORF29p nucleic acids, and a monoclonal antibody against the ORF29p protein. Compositions of the invention are useful for the delivery of a prophylactic or therapeutic agent to a eukaryotic cell, particularly a human cell. Examples of agents that can be delivered to a cell include proteins such as insulin, factor VIII, factor IX, and proteases; polynucleotides (e.g., for use in gene therapy), and organic compounds such as vitamins and a wide variety of pharmaceuticals (e.g., antineoplastic, antidiabetic or immunosuppressive drugs).

Sequence 1203 AA;

Query Match 100.0%; Score 6294; DB 22; Length 1203;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MENTQKTVTPTGLGVYACRVEDDLEISFLAARSTDSIALPLMNLVETKFTFS 60
 1 MENTQKTVTPTGLGVYACRVEDDLEISFLAARSTDSIALPLMNLVETKFTFS 60
 61 SLAVSGARTTGLAGITLKLTTSHFYPVSVFPHGKHVLPSSAAPNLTRACNAARERF 120
 61 SLAVSGARTTGLAGITLKLTTSHFYPVSVFPHGKHVLPSSAAPNLTRACNAARERF 120
 121 GFSCQPPVDGAVETTGAIICITRLGLEPENTILYLVTALFKEAVFMCNVFLHYGGLDI 180
 121 GFSCQPPVDGAVETTGAIICITRLGLEPENTILYLVTALFKEAVFMCNVFLHYGGLDI 180
 181 VHINHGDIIRIPFPVQLFMDVNLVPDPFNTHRSIGEGFYVPTFYNTGLCHLHDC 240
 181 VHINHGDIIRIPFPVQLFMDVNLVPDPFNTHRSIGEGFYVPTFYNTGLCHLHDC 240
 241 VIAPMAVALVRNVTAARGAAHLAFDENHGAVALPFDITYFYQSSSGTTTARGARN 300
 241 VIAPMAVALVRNVTAARGAAHLAFDENHGAVALPFDITYFYQSSSGTTTARGARN 300
 301 DVNSTKPSGSGFERRLASIMAAADTALHAETVNTGIYEETPTDIKEWPMFIGEGTLP 360
 301 DVNSTKPSGSGFERRLASIMAAADTALHAETVNTGIYEETPTDIKEWPMFIGEGTLP 360
 361 RLNALGSYARVAGVIGAMVFSNSALYLTEVEDSGMTEAKDGGPSPNRFYQAGPHL 420
 361 RLNALGSYARVAGVIGAMVFSNSALYLTEVEDSGMTEAKDGGPSPNRFYQAGPHL 420
 421 AANPQTRDGHVLSSTGSSNTFESVDYLAICGFGAPLLARLLFYLERCDAGFTGGH 480
 421 AANPQTRDGHVLSSTGSSNTFESVDYLAICGFGAPLLARLLFYLERCDAGFTGGH 480
 481 GDALKVTGTFTDSEIPCSI CEKHTFPVCAHTVHLRQMRPFGQATPQIGVFTWNSQ 540
 481 GDALKVTGTFTDSEIPCSI CEKHTFPVCAHTVHLRQMRPFGQATPQIGVFTWNSQ 540
 541 YSDCDPLGNVAPVILIRKPGDQTEAAKATMDTYRATLERFIDLEQRLLDRGAPCSSE 600
 541 YSDCDPLGNVAPVILIRKPGDQTEAAKATMDTYRATLERFIDLEQRLLDRGAPCSSE 600

601 GLSSVIVDPTFRRLDILRLARIEQITTOFMKLVYETDYKIREGISEATHSMALTDPY 660
 601 GLSSVIVDPTFRRLDILRLARIEQITTOFMKLVYETDYKIREGISEATHSMALTDPY 660
 661 SGAFCDITNFKVRLHVAVVDLALSQCHCVFYGOQVEGRNFRNQFQVLRFRFVLDLFG 720
 661 SGAFCDITNFKVRLHVAVVDLALSQCHCVFYGOQVEGRNFRNQFQVLRFRFVLDLFG 720
 721 GFSTSTSIIVTISEGPVSAPNLTGODAPAGTTPGDLARVSEVIRDIRVKNRVVFSGN 780
 721 GFSTSTSIIVTISEGPVSAPNLTGODAPAGTTPGDLARVSEVIRDIRVKNRVVFSGN 780
 781 CTNLSEARARLVGLASAYQROEKVDMLHGLGFLKQFHGLLFPFGMPNPKSPNPOW 840
 781 CTNLSEARARLVGLASAYQROEKVDMLHGLGFLKQFHGLLFPFGMPNPKSPNPOW 840
 841 FWTLLQRNOMPADKLTHEITIAAVKFTTEYAAINFINLPPTCIGELAQFYMANLILK 900
 841 FWTLLQRNOMPADKLTHEITIAAVKFTTEYAAINFINLPPTCIGELAQFYMANLILK 900
 901 YCDHSQYLINTLTSITGARRPRDPSSVLHWRKDVTSAAADIETQAKALLEKTENLPELW 960
 901 YCDHSQYLINTLTSITGARRPRDPSSVLHWRKDVTSAAADIETQAKALLEKTENLPELW 960
 961 TTAFTSTHLVRAAMNQRPNVVLGISISKYHGAAGNRRVFAQGNWSGLNGGKVCPLTFD 1020
 961 TTAFTSTHLVRAAMNQRPNVVLGISISKYHGAAGNRRVFAQGNWSGLNGGKVCPLTFD 1020
 1021 RTRRTIACPRGGFTCPVTGSSGNRETTLSDOVRGIIIVSGAMVOLAIYATVVRVARG 1080
 1021 RTRRTIACPRGGFTCPVTGSSGNRETTLSDOVRGIIIVSGAMVOLAIYATVVRVARG 1080
 1081 AQHMAFDWLSLTDDEFLARDLEELHDOIIQTLETPTWTEGALEAVKILDEKTTAGGET 1140
 1081 AQHMAFDWLSLTDDEFLARDLEELHDOIIQTLETPTWTEGALEAVKILDEKTTAGGET 1140
 1141 PTNLAFNPDSCPSHDTTSNVLNIGSNISGSTVPGIKRPPEDDELFDLSGIFIKHGNIT 1200
 1141 PTNLAFNPDSCPSHDTTSNVLNIGSNISGSTVPGIKRPPEDDELFDLSGIFIKHGNIT 1200
 1201 MEM 1203
 1201 MEM 1203
 RESULT 2
 AAE25598
 ID AAE25598 standard; Protein; 1452 AA.
 XX AAE25598;
 AC
 XX
 XX
 DT 04-NOV-2002 (first entry)
 XX
 DE Herpesvirus ICP8-GFP fusion protein.
 XX
 KW ICP8 protein; herpesvirus; green fluorescent protein; GFP; infection;
 KW antiviral agent; fusion protein.
 XX
 OS Chimeric - Herpesvirus.
 OS Chimeric - Unidentified.
 XX
 XX US6399354-B1.
 XX
 XX 04-JUN-2002.
 XX
 XX 31-JUL-1998; 98US-0127227.
 XX
 XX 31-JUL-1998; 98US-0127227.
 PR (HARD) HARVARD COLLEGE.
 XX
 XX Knipe DM, Taylor TJ, McNamee EE;

PT viral protein derived from virus and detectable protein
 XX
 PS Disclosure; Fig 7; 30pp; English.

CC The invention relates to a fusion protein comprising a viral protein with
 CC at least one function and a detectable protein, fused to maintain the
 CC open reading frame. The fusion protein is useful for determining whether
 CC a cell is a virus-resistant cell or a virus-susceptible cell, by
 CC contacting a virus that expresses the fusion protein with the cell to be
 CC tested, under conditions sufficient for the virus to infect the cell, and
 CC detecting the presence or absence of the fusion protein, where the
 CC absence of the fusion protein identifies a virus-resistant cell and the
 CC presence of the fusion protein identifies a virus-susceptible cell. The
 CC fusion protein is useful for identifying an anti-viral agent or an agent
 CC that blocks the expression of the fusion protein by contacting a virus
 CC that expresses the protein, a host cell and the agent to be tested, in
 CC conditions sufficient to allow for the virus to infect the cell, and
 CC determining the amount of the virus present, where a decrease in the amount
 CC of virus present identifies the agent. This sequence represents an
 CC ICP8-green fluorescent protein (ICP8-GFP) fusion protein of the
 CC invention.

XX Sequence 1452 AA;

Query March 48.5%; Score 3054; DB 24; Length 1452;
 Best Local Similarity 49.8%; Pred. No. 2.5e-288;
 Matches 609; Conservative 200; Mismatches 366; Indels 48; Gaps 17;

1 MENTQKWT--VPTGPIGVY--ACRVEDLDLEISFLAARSTDSNALLPLMRNLTVS 55
 1 METKPTATIKVPPGPIGVYACRSEGL--LALLSARGDADVAAPLVWGLTVS 58
 56 KTFSSIAVSGARTTLAGAGITLKLTHSHFVPSVFVFGHGHVLPSSAAPMLTRACNA 115
 59 SGFEANVAVVGGRTTGLGTAVSLKTPSHYSSVVVFGHGHLPSTQAPNLTCLR 118
 116 ARERFGRSCGPPVDGAVTTGAEICTRLGLSPENTILYLVTALFKAVFVWCVNLFHY 175
 119 ARRHFGSDVTPRGGDKHETTCALCERGLGDPDRALLYLVTVEGKEAVCINNFTLH 178
 176 GGLDIVHINHGDIPLFPVQLFMPDVRVLPDPFNTHRSIGEGVYPTPTNYTGLCH 235
 179 GSGDKVTIGAEVHRIPVYQLFMPDFSEVIAEPENANHSIGENTYPLFPFNPTR 239
 236 LIHDCVTAPMAVALRVNTAVAGAAHLAFDENHGVLPDPITVYFQSSSOTTAR 295
 239 LLEAVVGPAAVALRCRNVDAVARAAHLAFDENHGAALPADITTAFAEQG--KTPR 296
 296 GARNDVNSTSKPSGSGFERRLASIMADTALHAEVIENTGIVETPTDIKEMPFICM 355
 297 GGR-----DGGKGPAGGEORLASVWAGDAALALESIIVSMVDFDPPDTSAPLCEQ 351
 356 EGTLPRLNAGSTAVAGVIGAMVSPNSALYLTVEDESGMTEAKDGGFSPNRYQF 415
 352 DTAARANAVGAYLARAAGLVGAMVSTNSALHLTEVDAGPADPKDHSK--SFYRFLV 410
 416 AGPFLAANPOTDRGHVL-----SSQTSNGSTNPSVDYLALICGAPLALLPYL 468
 411 PGTHVAAAPQVDEGHVVPFGFGRAPFAPVGGTQ--EFAGEHLAMLCGFSAPALLPYL 469
 469 ERCDAGFTGGHG--DALKYVTGTFDSEI PCSLCEKHTRPVCAHTTVHRLQRMFPGQAT 527
 470 ERCDGGVIGRQEMDFRVVADSNQTDVPCNLCTEDTRHACVHTTLMRLRAPHFASAA 529
 528 ROPIGVGTWNSOYSCDPLGNVAPYLIIRKGGDQTEAAKATWQDTYRATLERLFDLEQ 587
 530 RGAIGVGTWNSOYSCDPLGNVAPYLIIRKGGDQTEAAKATWQDTYRATLERLFDLEQ 587
 588 ERLLDRGAPCSSLSSVVDHETFRILTRAEQTTTFQFMVLTETRYDKIREGLS 647
 589 LQVVDQAVPAMGLTETITNEALHTVNNVQVVDREVEQLMNLVEGRNFKFRDGLG 648
 648 EATHSMALTFDPSYGFAPCIPITFLVLRTHLAVVQDIALSQCHVCYGOQVEGRNFRNQF 707

649 EAHAVSTLDPYACGCPCLLQLLGRSNLAVYQDIALSQCHGVFAGOSVEGRNFRNQF 708
 708 PVLRERFVLDLFGGFGISTRSITVTLSEG--PVSAFNPTIGDAPAGRTFDGDLARVSEVI 766
 709 PVLRERVDMFNNGFSLAKITLVALSEGAACIAPSLAQGTAPAESFEGDVARTVIGFP 768
 767 RDIRVNRVPSGNCNLSAARARLVGLASAYQRQKRVDMHGLGFLKQFPHGLFP 826
 769 KELRVKSVLFGASASANAASEAKARVASLQAYQKPKGVNDILGLPLGLLKQFAAIFP 828
 827 RCMPPNSKSNPQWFTLLOKNOWPADKLTHERITTAAYKRFTEEVAAINFINPTCI 886
 829 NGKPPGSGNQPPQWFTALQKQPPARILSFEDLETAFIKGFELDYGAINFINLAPNV 888
 887 GELAFYMANILKYCDHSOYLINTLSITGARRPRDPSVHLWIKRDKVTSADITQA 946
 889 SELAMVYMANQILRYCDHSTYFINTLITAGSRPSVQAAAAW---SAQGGAGLEAGA 945
 947 KALLEKTENLDELWTTAFTSTHLVRAANNQRPVVLGISIKYHGAAGNRVFOAGNWSG 1006
 946 RALMDAVDAHPGAWTSMFASCNLLRFVMAARPMVVLGLSISKYGMAGNDRVFOAGNWS 1005
 1007 LMGGNVVCPLFTFDRTRRFIACPRGCFICPVTPGSSGNRETTLSDOVRGIIVSGGAMVQ 1066
 1006 LMGKNACPLLIIFTRTKFVLACPRAGFVCAASNLGGAAHSSICEQLRGIIEGGAAVA 1065
 1067 LAIVATVVRVAGARQAMAFDDMLSLTDDDFLARDLELHDQIITQLETPTWVEGAL--- 1123
 1066 SSVFVATVKSIGPPTQQLQIEDMLALEDEYVLSSENMELTARALRGNGSWSTDALEVA 1125
 1124 -EAVKILDEKTTAGDGETPNTLAFNED--SCEPSHDTTSNVLNISGNSISGTVPGIKRP 1180
 1126 HEBAALYSQGNAGE-----VFNFGDFGCB-----DDNATPFGGAGPAPAFGRKKA 1173
 1181 PEDDELFDLGGIPKKGNTWEM 1203
 1174 FHGDDPFG-EGPPDKGDLTDM 1195
 RESULT 4
 ID AAW72105 standard; Protein; 1196 AA.
 AC AAW72105;
 XX
 XX 18-DEC-1998 (first entry)
 XX HSV-2 strain SB5 Contig ID 12 ORF#2 protein.
 XX HSV-2 strain SB5; immunological response induction; therapy;
 XX antiviral identification; viral protein inhibitor.
 XX Herpes simplex virus type 2.
 XX W08820016-A1.
 XX 14-MAY-1998.
 XX 31-OCT-1997; 97MO-US20016.
 XX 09-JUN-1997; 97US-0049018.
 XX 04-NOV-1996; 96US-0030279.
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 XX Chan JY, Dabrowski-Amarel CE, Delvecchio AM, Dillon SB;
 XX Esser KM, Leary JJ;
 XX WPT; 1998-286847/25.
 XX N-PSDB; AAV62156.
 XX Herpes simplex virus type-2 sequences - useful in, e.g. prevention

PT and treatment of infection or inducing immunological response in
XX mammal
PS Claim 10; Page 83-84; 748pp; English.
XX This sequence represents a Herpes simplex virus type-2 (HSV-2) protein
CC sequence of the invention. This sequence was isolated from a HSV-2 strain
CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 12.
CC The proteins can be used for the treatment or prevention of disease, to
CC induce an immunological response in a mammal or to identify inhibitors,
CC activators or novel antivirals. Antagonists of the proteins can be used
CC to inhibit a viral polypeptide. The DNA sequence or a vector containing
CC it can also be used to induce an immunological response in a mammal.
XX Sequence 1196 AA;
Query Match 47.9%; Score 3016.5; DB 19; Length 1196;
Best Local Similarity 49.1%; Pred. No. 8.4e-285;
Matches 596; Conservative 211; Mismatches 367; Indels 41; Gaps 16;
Y 4 TQKTVPTGPGYVY--ACRVEDLDLEISLAARSTDSLDLALPLMRLNLTVEKTFSS 61
b 7 TTTTVKVPGPYVYGRACPAEGL--LSLSARSADAVAVAPLIVGLTVSGFEAN 64
Y 62 LAVSGARTTLAGAGITLTKTSHFYPSVVFVHGKHLVPSAAPNLTRACNAARERFG 121
b 65 VAAVVGSRITGLGGTAVSLKMPHSYVSVYVFGHRLAPSTQAPNLTRLCERARRHF 124
Y 122 FSRGCGPPVGGAVETGABICRTRGLEBENTILYLVVLTALPKAVFMCNVFLHYGGLDIV 181
b 125 FSDYAPRPCDLKHETGTGALCERLGLDPRALLYLIVTEGFREAVCIISNTFLHGGMDKV 184
Y 182 HINHGDEVIRIPLFPVQLEMPDVNRVLPDPFNTHRSIGEGFYPTFPVNTGLCHLHDCV 241
b 185 TIGDAEVHRIPVYPLQMPDPFSRVIAFPNCNHSIGENFYLPFNRLPLALLFEAV 244
Y 242 IAPMAVALRVNVTAVAGAAHLAPDENHEGAVLPDITYTYFSSSGSTTTARGARRND 301
b 245 VGPAVALRARNVDVARAAHLAPDENHEGAAFPADITTAFAESQ--KPGSGAR-- 299
Y 302 VNSTSKPSPSGFERELASIMAAITALKAEVIFNTGIYEEPTDIKEWPMFIGNETLPR 361
b 300 --DAGNKPGAGFEGORLASWAGDAALESIVSAVDEPPDITTWLLEGGQETPAAR 357
Y 362 LNALGSYTRAVAGVIGAMVFNAGLYLTYEVDGSMTEAKDGGPGPSFNRFYQAGPHLA 421
b 358 AGAVGAYLARAAGLVGAMVFNAGLYLTYEVDGSMTEAKDGGPGPSFNRFYQAGPHLA 416
Y 422 ANPQTDREGHV-----SSQSGSSNTEFSVDYLAICGFGAPLLARLLFYLERCDAG 474
b 417 ANPQDREGHVVGVEGRTAPLVGGTQ--EPAGEHFLAMLCGFSAPALLAKMLFYLERCDGG 475
Y 475 AFTGGHG--DALKYVTGTFDSIPCSLCKHTRPCVCAHTTVHRLRQMPRFGQATROPIGV 533
b 476 VIVGQENDVRYVADSQOTVPCNLCTFETHACHTTLMRLRHPKFAASAARGAIGV 535
Y 534 FGTWNSQVSDCDPIGNVAPYILRLKPGQTEAAKATMDTYRATLERFLDLRQERLLDR 593
b 536 FGTWNSAYSDCDVLGNVAAFAALKE--ADGSENTRTIMQETVRAATERVMALEALQVQD 594
Y 594 GAPCSSEGLSVIVDHPFRILDTLRAIREQTTFQMKVLVETRDYKIREGLSEATHSM 653
b 595 AVPTALGRLEITIGTRREALHTVNNIKQLVDREVEQLMRNLIEGRNFKFRDGLAEANHAM 654
Y 654 ALTPDPYSGAFCPITNLFVRLKTHLAVVQDLALSCHVYFGQVQVQVGNFRNCPQVLR 713
b 655 SLSDPVTGCPPLQLLARRSNLAVVQDLALSCHVYFGQVQVQVGNFRNCPQVLR 714
Y 714 FVDLFGNGFTSTRITVTLSEG--PVSPNPTLGDAPAGTFCGDLARVSVEVIRDIRVK 772
b 715 VMDLFNNGFLSAKTLTVALSEGAIAPSLTAGQTAPESSFEGDVARVTLGFPKELRVK 774
Y 773 NRVVFSNGCTNLSEARARLVGLASAYQORKEKRVMDLHGALGFLLLQFHGLLPRGMPN 832

Db 775 SRVLFAGASANAEEAAKARVASLSQSAVQKPRVDILLGLFLGLKQFHAVIFENKPG 834
QY 833 SKSPNCPWFLLQRNCPADKLTHEITIAAVKPTTEYAAINFINLPPTCIGELAQF 892
Db 835 SNQPNQWFTALQRNCPARLLSREDIETIAFKFSLDYGAINFILAPNNVSELAMY 894
QY 893 YMANILKYCDHSQYLINTLTSITGARRPRDPSVHLWIRKDVTSAAADIETCAKALLEK 952
Db 895 YMANQILRYCDHSTYFINTLTVIAGSRPPGVQAAAAPQ---CGAGLEAGARALMDS 951
QY 953 TENLPELWTFATSTHLVRAAMQRMVVLGISISIKYHGAAGNNRVFOAGNWSGLNGKN 1012
Db 952 LDHPGQAWTSMFASCNLLRPVWAAREPMVVLGLSISIKYHGAAGNNRVFOAGNWSGLNGKN 1011
QY 1013 VCPLEFTRTRRRIIACPRGGFCIPVTPSSGNRETTLSDOVRGIIVSGAMVQALAIYAT 1072
Db 1012 ACPILLFDRTRKVLACPRAGFVCAASSLGGGAHSLCEQLRGIIAEGGAIVASSVFA 1071
QY 1073 VTRAVGARQHMADFDDWLSLTDDEFLARDLEBLHQIIOETLETPMTVEGAL---EAVKI 1128
Db 1072 TVKSLGPRTOQLQIEDWLALLEDEVLSSEMEFTTRALERGHGEWSTDAALVAAHEAEL 1131
QY 1129 LDEKTTAGDGETPTNLAFNDFSCERSHDTTSNVILNIGSNISSGTVPGIKRPPDELEFD 1188
Db 1132 VSQGAAGE-----VFNFGDGEDDDHAASFGGLAAA--AGRAGVARKRAFHGDDPFG 1182
QY 1189 LSGIPIKHGNTMEM 1203
Db 1183 -EGPPEKK-DLTUDM 1195
RESULT 5
AAW72197
ID AAW72197 standard; Protein; 623 AA.
XX AAW72197;
XX 13-JAN-1999 (first entry)
DT
XX HSV-2 strain SB5 Contig ID 15 ORF#32 protein.
DE HSV-2 strain SB5; immunological response induction; therapy;
XX antiviral identification; viral protein inhibitor.
KW
XX Herpes simplex virus type 2.
OS
XX W09820016-A1.
PN
XX 14-MAY-1998.
PD
XX 31-OCT-1997; 97WO-US20016.
PF
XX 09-JUN-1997; 97US-0049018.
PR
XX 04-NOV-1996; 96US-0030279.
PA (SMIK) SMITHKLINE BEECHAM CORP.
XX Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB;
PI Esser KV, Leary JJ;
XX WPI; 1998-286847/25.
PI N-PSDB; AAV62176.
XX
XX Herpes simplex virus type-2 sequences useful in, e.g. prevention
PT and treatment of infection or inducing immunological response in
PT mammal
XX Claim 10; Page 123; 748pp; English.
PS This sequence represents a Herpes simplex virus type-2 (HSV-2) protein
CC sequence of the invention. This sequence was isolated from a HSV-2 strain
CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.

Db 478 LVSQGAAGE-----VFNGDFGDEDDHAASFGGLAA--AGAGVARKAFHGDOPF 528

2y 1188 DLSGIPKIGNITMEM 1203

Db 529 G-EGPPEKK-DITLDM 542

RESULT 7

AB53126

AB53126 standard; Protein; 1132 AA.

AB53126;

28-FEB-2001 (first entry)

Macaca mulatta rhadinovirus 17577 RRV ORF6 protein SEQ ID NO:9.

Macaca mulatta rhadinovirus 17577; RRV; rhesus macaque rhadinovirus; Genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6; IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine; cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma; lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy; splenomegaly; hypergammaglobulinemia; autoimmune haemolytic anaemia.

Macaca mulatta rhadinovirus 17577.

W0200028040-A2.

18-MAY-2000.

05-NOV-1999; 99WC-US26260.

06-NOV-1998; 98US-0107507.

20-NOV-1998; 98US-0109409.

(UYOR-) UNIV OREGON HEALTH SCI.

Wong SW, Axthelm MK, Searles RP;

WPI; 2000-376552/32.

New rhesus rhadino virus for producing non-human primate model useful for testing potential treatments and efficacy of the candidate vaccine for conditions associated with RRV infection

Claim 5; Page 123; 141pp; English.

The present invention describes a novel rhesus macaque rhadinovirus called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the RRV genome sequence, and AAB53204 represents the proteins encoded by the genome sequence. The present invention also specifically claims the individual open reading frame (ORF) nucleotide sequences from the genome which encode the individual proteins, but these sequences are not given. A non-human animal infected with RRV can be used for testing the efficacy of drug in the treatment of condition associated with infection with RRV such as Kaposi's sarcoma, lymphoproliferative disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly, hypergammaglobulinemia or autoimmune haemolytic anaemia, by administering the drug to a immuno-compromised non-human primate preferably Rhesus macaque monkey obtained by as a result of infection by Simian Immunodeficiency Virus (SIV). RRV is useful for producing non-human primate model for testing potential treatments for conditions associated with RRV infection. It is also useful for testing the efficacy of the candidate vaccine against RRV infection or conditions associated with its infection by administering the vaccine to the subject capable of infection with RRV, inoculating the subject with RRV and observing the effect of vaccine. AAC64755 to AAC64765 and AAB53205 to AAB53213 represent sequence used in the exemplification of the present invention.

Sequence 1132 AA;

Query Match 9.9%; Score 620; DB 21; Length 1132;

Best Local Similarity 22.6%; Pred. No. 2.7e-50;

Matches 290; Conservative 212; Mismatches 506; Indels 274; Gaps 50;

Qy 1 MENTOKTVPTGTLGYVYACRVEDLDLEISPLAARSTDSLDLPLMRNLVTEKTFIS 60

Db 11 LEDNQS-RAPICAGCYVAYSKODFFFAEASTLGNRPSGSGVFSPLILYGLTVEHEFFL 69

Qy 61 SL-AVMSGARTTGLAGAGITLKLTTSHFYPSVVPFHGKHVLPSSAAPNLTRACNAAR 119

Db 70 TVKAAVKKVDTTTTLA-----VKYTC--FHREIVFHNASLFRPVPDGTGLNELCEARAL 122

Qy 120 FGSRC--QOPP--VDGAVETTGAEICTRIGLEPENTILYLVVTLFKE-----AV 166

Db 123 FGYQFIEPGPHSIWNPLE-----CPQLP-DKDEMFLGVVVTEGFKERLNRGCLVPAV 175

Qy 167 FMCNVFLHYGGLDIVHINEGDVIRIPLFPVQLFMPDVNRLVDPDPNTHERSIGEGVYPT 226

Db 176 FQYQQ-----VQIAGQAFKFLYDEDFAPGHNM-----P 207

Qy 227 PFYNTGLCHLIHDCVIAPMAVALRVNVTAVARGAAHLAFDENHEGAVLPDITVTFOS 286

Db 208 RFYHKDVSAYLVDSLFTSIAQALRLKDVAVIHATEKQPMODHYKIAKI---VQAKQF-- 262

Qy 287 SSGGTTTARGARENDVNSTKSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDI 346

Db 263 -----STLPKTTDGGSHMTVDVSVAEALSYGCMF-----LSCPQDA 300

Qy 347 KE-----WPMFIGMEGLTFLRNALGSYTVAVIGVIGAMVFPNSALYTEVEDSGMTEA 400

Db 301 CELLNYDSWPIFDGCDSPPEARVNALERNSAEQAVVAGOLPAANSVLVLTQVKQAPRQ 360

Qy 401 KDGPGPSFNRYQFAGPHLAANPQTDGRH-----VLSQSTGSSNTEFSVDYIALICG 455

Db 361 K-GDVVYNSFFLOHGLFLNEATIKENGSAFKGVPSNALDGSTFYP---HLAYAAS 415

Qy 456 FGAPILARLLFYLERCDAGFTGGHGDALKVVTGTFDSEIPCSCLEKTRPVCATTVHR 515

Db 416 FSPHLLAKLYNQFLOHKKSTNQAFNMVHVGTAAENSEMCTLCHGNTPATCLNTLYR 475

Qy 516 LRQWRPFQATQ-PIGVFGTMNSQYDCDPLGNYAPYLILRKFQDOT-EAAKATMODT 573

Db 476 LKDRPFAVTPQRRDPYVVTGTAGT-FNDLEILGNFASFRDREEDGNPADEHPKYVYQ 534

Qy 574 YRATLERFLDLQERLLDRGAPCSSEGLS-----SVIVDHTFRILTLRARIQT 626

Db 535 CQVTEKL-----SAIGITEDHNVNLTNIQSFLAVFKGDSIVDGE 578

Qy 627 TTQFMKVLVETRYKIREGLSEATHSMALTDPYSGAFCPITNPLVXRLHVLAVVDLALS 686

Db 579 VMKFNYSMIK-NNFNREHVKSVHILQFCNVYQWQAPCAVFLNLYKSLWIIQDCLP 637

Qy 687 QCHCVYQO-----QVEGNFRN-----QFQPVLRFRFVDLF 718

Db 638 --YCMYEQDNPAMGILPSEWLKHQFQTLWTFKAACLDLGRVLTGCELKIVHRDMFCDF 695

Qy 719 -----NGQFISTRTITVLSGPNAPNPTLQDAPAGRTPDGLARVSVEVIEDIRVN 773

Db 696 DTDAGSNGMLAPFKQVYR-----IARMMVVPKSIKKN 729

Qy 774 RVVFGNCTNLSAARAPLV--GLASAYQKQKRVDMHGLGFLKQPHGLLFPFGMP 830

Db 730 RIIFSN--TAGSEAVQSGFVXPTGTRTY-----VYAGPYMKPLNSLHRALE----- 774

Qy 831 PMSKSPNQWMTLLQONMPADK-LTHEETITIA-VKFTSEYAININLPPTIGE 888

Db 775 PDTKTAALYHKKISQNKTVLKDVPDDDELAELVSVKTNLSLAFTETNVDVVDLSMS 834

Qy 889 LAQFYMANLILKYCDHSQYLINTITSITGARPRDPSSVLHMRKDVTSAAADIETOKA 948

Db 835 YARIKNGAILRACQIQFYATTL-HCLTPVLOIDAEEYPH-----VLGSAIATPVAY 888

Qy 949 LLE---KTENLPETWTATFTSTHLYRAMNQRPVVLGISIKSHGAGNNRVFOAGN-- 1003

Db 889 LAEIRGT-----ALTQTARQPVAAGRLRPVITVPMVKNKTVGNVNNVHFCNGLG 943
 QY 1004 -WSGLNGKNVCPLFT-FDRT-----RRFIACPRGGFICPVTPGSSGNRETTLSQDV 1054
 Db 944 YFAGGVDRNLWPSSPPFKTGVSAMLRKRHVMTTP-----IIDRLIKRAAG--QTISTFEA 998
 QY 1055 RGIIVSGGAMVQ-----LAIYATVVRVAVARAQHAQFDDWLSLTDDEFLARDLEELH 1106
 Db 999 ESKVRSVQALLKEDKNPNLLKSVILELRHILKGCQD-----LSEEDVQYLYGDYCMLT 1052
 QY 1107 DQIQTGLE-----TPWTVEGALEAVKILDKTAGDGETPTNLAFNFDSCPESHD---T 1157
 Db 1053 DEVLFTLDNIAQSGVFWTIE---DAGALIEDQDADLQFVDSDDIATASCQPEEQLEPT 1109
 QY 1158 TSNVLNIGSGNISGTVPLKX 1179
 Db 1110 PS-----AGALLACKX 1121

RESULT 8
 AAY26935
 ID AAY26935 standard; peptide; 35 AA.

AC AAY26935;
 XX
 DT 21-DEC-1999 (first entry)
 DE Varicella zoster virus type 29 recombinase D35E motif.
 KW Recombinogenic motif; retrovirus; integrase; invertebrate; transposase;
 KW consensus; vaccine; gene therapy; allergy; autoimmune disease; cancer;
 KW cardiovascular disease; graft rejection; hematopoietic disorder;
 KW immunosuppressive disorder; immunoproliferative disease; jaundice;
 KW immunodeficiency disease; infectious disease; inflammatory disease;
 KW septic shock; metabolic defect.
 XX
 OS Varicella zoster virus.
 XX
 PN US5959074-A.
 XX
 PD 28-SEP-1999.
 XX
 PF 28-FEB-1997; 97US-0807332.
 XX
 PR 01-MAR-1996; 96US-0012616.
 PR 02-AUG-1996; 96US-0023064.
 XX
 PA (NAJE-) NAT JEWISH MEDICAL & RES CENT.
 XX
 PI Gelfand EW, Dreyfus DH;
 XX
 DR WPI; 1999-561074/47.
 XX

Peptides derived from a Herpes virus recombinogenic motif useful for the
 production of vaccines and in gene therapy -
 Claim 1; Column 37-38; 29pp; English.

The invention relates to novel recombinogenic motifs found in
 (retro)viral integrases which have similarity to invertebrate
 transposase molecules. The peptides, or their homologues, have the
 following identifying characteristics: (a) an initial and terminal amino
 acid comprising Asp or Glu, separated by 34-39 amino acids, especially
 35 amino acids (the D35E motif); (b) a high probability ($p < 0.05$) of
 alignment with a consensus sequence (AAY26941), as determined by:
 (1) aligning the initial and terminal residues of the homologue sequence
 with those of the consensus; (2) aligning the intervening residues of
 the homologue with those of the consensus by maintaining the spacing of
 the consensus while, if necessary, altering (by the insertion of spaces
 or the deletion of residues) the spacing of the homologue; and
 (3) generating a distribution of 1500 random sequences of amino acids
 between the initial and terminal residues and identifying the percentage
 of aligned residues. The recombinogenic motifs may be useful in the

CC production of vaccines and in gene therapy to treat allergies,
 CC autoimmune diseases, cancers, cardiovascular diseases, graft rejection,
 CC hematopoietic disorders, immunosuppressive disorders,
 CC immunoproliferative diseases, immunodeficiency diseases, infectious
 CC diseases, inflammatory diseases, jaundice, septic shock and other
 CC immunological, genetic or metabolic defects.

XX Sequence 35 AA;

Query Match 2.9%; Score 183; DB 20; Length 35;
 Best Local Similarity 100.0%; Pred No. 3.5e-10;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 545 DPLGNYPYLIRKPGDQTEAAKATMDTYRATLE 579
 Db 1 DPLGNYPYLIRKPGDQTEAAKATMDTYRATLE 35

RESULT 9

AAB73931
 ID AAB73931 standard; peptide; 35 AA.

XX
 AC AAB73931;
 XX
 DT 29-MAY-2001 (first entry)
 DE Varicella zoster VZV29 protein recombinogenic sequence.
 XX
 KW Varicella zoster virus; VZV29; herpesvirus; D35E; anti-viral;
 KW immunosuppressant; immunostimulant; cytostatic; anti-allergic; cardiant;
 KW anti-inflammation; gene therapy; infection; herpes virus replication;
 KW V(D)J recombination; retroviral integrase;
 KW immunoglobulin class switching; allergy; autoimmune disease; cancer;
 KW cardiovascular disease; graft rejection; haematopoietic disorder;
 KW immunological disorder.
 XX
 OS Varicella-zoster virus.
 XX
 PN US6187584-B1.
 XX
 PD 13-FEB-2001.
 XX
 PF 23-JUN-1999; 99US-0338876.
 XX
 PR 01-MAR-1996; 96US-0012616.
 PR 02-AUG-1996; 96US-0023064.
 PR 28-FEB-1997; 97US-0807332.
 XX
 PA (NAJE-) NAT JEWISH MEDICAL & RES CENT.
 XX
 PI Dreyfus DH, Gelfand EW;
 XX
 DR WPI; 2001-256131/26.

New nucleic acid molecules encoding peptides with transposase
 activities, useful as immunosuppressants, anti-viral agents, or vectors
 for gene therapy of diseases, e.g. allergies, autoimmune diseases,
 cancers or graft rejection -
 Claim 1; Column 25; 31pp; English.

The present sequence is provided in a specification relating to a new
 isolated nucleic acid molecule comprising a nucleic acid sequence
 encoding a peptide consisting of one of ten amino acid sequences
 defined in the specification or homologues of these sequences.
 The nucleic acid molecules are useful for producing peptides.
 Important in the regulation and function of herpes virus replication,
 V(D)J recombination, the retroviral integrase function or immunoglobulin
 class switching. The isolated nucleic acid molecules and the peptides
 they encode are useful as immunosuppressants, anti-viral agents or
 vectors for gene therapy. These are particularly useful in protecting
 humans from a variety of diseases, e.g. allergies, autoimmune diseases,
 cancers, cardiovascular diseases, graft rejection, haematopoietic

CC disorders, immunodeficiency diseases, immunoproliferative diseases,
CC immunosuppressive disorders, infectious diseases, inflammatory diseases,
CC jaundice, septic shock, or other immunological defects, as well as other
CC genetic or metabolic defects.

SQ Sequence 35 AA;

Query Match: 2.9%; Score: 183; DB: 22; Length: 35;
 Best Local Similarity: 100.0%; Pred. No.: 3.5e-10;
 Matches: 35; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;

545 DPLGNVAPYLILRRPGDQTEAAKATMQDTYRATLE 579
 1 DPLGNVAPYLILRRPGDQTEAAKATMQDTYRATLE 35

RESULT 10

AAB07562
ID AAB07562 standard; Protein; 1841 AA.
...

AAB07562:

20-OCT-2000 (first entry)

Protein encoded by the bleomycin (BLM) gene cluster ORF30.

W BLM gene cluster; bleomycin gene cluster; polyketide metabolite;
W bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;
W thiazoline; bithiazoline; microbial metabolite; sugar
Y

5 Streptomyces verticillus.

N WO200040704-A1.

X D 13-III.-2000

✕

00 CAF-2000; 2000WU-US00445.

R 06-JAN-1999; 99US-0115435.
R 05-FEB-1999. 99US-0118848

R 05-JAN-2000; 2000US-0477962.

A (REGC) UNIV CALIFORNIA.

Shen B, Du L, Sanchez C, Chen M, Edwards DJ;

R WPI; 2000-465974/40.

R X
N-PSDB; AAA58471.

New bleomycin gene cluster components useful for peptide and/or polyketide metabolites, especially bleomycin, production and for chemically modifying biological molecules -

Disclosure; Page 97-137; 162pp; English.

AA07556-78 represent proteins encoded by open reading frames (ORFs) 8 to 30 of the BLM (Bleomycin) gene cluster. The proteins encoded by the gene cluster are useful for producing peptides and/or polypeptides, especially bleomycin or bleomycin analogues. They are also useful for chemically modifying biological molecules to produce branched methyl groups, and for coupling amino acids and fatty acids. They may be reacted with an apo-carrier protein and coenzyme A to produce a holo-carrier protein. The BLM gene cluster or coenzyme A domains can be used individually or collectively to produce thiocolidine, thiazoline, bithiazoline and bithiazoline-containing microbial metabolites. The BLM gene cluster may also be used to produce sugars.

Sequence 1841 AA;

Query Match	2.5%	Score 156.5;	DB 21;	Length 1841;
Best Local Similarity	19.6%;	Pred. No. 0.00015;		
Matches 268; Conservative	133;	Mismatches 470;	Indels 493;	

Matches 268; Conservative 133; Mismatches 470; Indels 493; Gaps 67;

QY	14	PLGYVYACRVEDLDLEEISFLAARST-DSDLALLFLMRNLTVKFTFTSSLAV-----	520
Db	461	PMPLVVSARTREALAEAVRDLLAANSAPETGRLADLAATLAGRAFFVYRAAVVCHDLPEA	520
QY	66	----SGAR--TTGLAGAGITLKTTSHFYSVFVFGGKHVLPSSNAPNLTRACNAARERF	120
Db	521	ARLJGGARGETALPGR-----BAVFLP-PCGOTLPDPTGRLGYADVPAFPAHF	567
QY	121	GFSCRQGPFDVGAVE---TTGABICTRLGLEPENTILYLVVTLALFKAEAVFCNVFLHYGG	177
Db	568	-----DACAEGFAPLGTDLHAALGAPADDI--RAAQPALFAVEYALARTLMDWGV	615
QY	178	LDIVHINH--GDVIRIPLFVQLFPDQVNLVDPDPNTHRSIGEGFYPTFPFVNTGLCH	235
Db	616	RPAAMLGHSIGSYVAATLAGV-LSUPDALTLVRARAEAOH-TMPPGMLAVLPTDPLRP	673
QY	236	LIHD-----CV-----IAPMAVALRVNVTAVARGAAHLADENHEGAV--	274
Db	674	LLPPEVEFSAFNAPGRCVGGPPFVAELRAELARRGVPAEELATAH-AP--HSAARVP	729
QY	275	-----LPDDITYYFOSSSG-----TTT-----ARGARR-2999	
Db	730	LLDGRGVLEGVLRPPRLRYV---SSLTGWDADAATTPAYMLAHLRRRVPADGLRRC	786
QY	300	NDVN-----STKPSPSGGFFERLAS-----IWAADTA--LHA-EVIFNTGIY-339	
Db	787	LDLGPVALVETGPRAGLTGLARRAGCPGPYTRCIAAPDEAASLTHAVLWRSGCAV	846
QY	340	-----ESTPTDIK-----EWPMEFIGME356	
Db	847	DWTAFRRCPRRTTPVGYPPFORVHRHMDADESEPTDLATRALRSLTDGDPPLAVDQR	906
QY	357	GTLPRINALGVSYARVAGVIGAMVFSNSALYL-TEVEDSGMTEAKDGGPSPNRFVQ	414
Db	907	PGLRTGLNLCALARD-----YLATGEASGVL-----PGHFRFLD943	
QY	415	FAGHIAANPQTRDGHVLSQSGSSNTEFSVDYLALICGFAPLLARLLFVLRCDAG	474
Db	944	YLRTLAASAAPAD-DAGTITAAEITAAH-----PFSGLVDLLRHCAQG985	
QY	475	-----AFTGGHDALKYVTGTDFDSEIPCSCLEKTRPVCANTVHRLRORM520	
Db	986	YPRALSTPGALDVLVYPAGSGDLRRITG-----EGTADHRAATGRL1022	
QY	521	PRFG-----QATRQPIGVFTGNSYSDCDPLGNVAPVYLILRKPGDQTEAAKATMOOT	573
Db	1027	TRLAGSLLDRLAAREPGRPLRVLEAGAG---AGSLTQALVTRAPG-----RLDY1073	
QY	574	YRATLERLFI-----DLEQRELLDRGAPCSSEGLS-----SVIVDHT611	
Db	1074	HATDSRHFVTALGREARRGLDFVRAVRVLDIARDPGEQFAGERFDVVCGLDVVHATPD	1133
QY	612	FRRIIDTLRARIEOTTQFMKVLVEYTRDYKIRGLESEATHSMALTDPYSGAFCPIT---	668
Db	1134	LRTLTGLHLSLWAPDGT--LALLET-----TADDPWLTIWGLTDGW1173	
QY	669	NFLVKRTH-----LAVQDLALSQCHVFCYGOQVEGRNFRNQ-----FOFVL710	
Db	1174	WHHTDRRTHGPLDLAGWRALLAGEDFATADVPPDGPQDAALLASQTPRPAAPASPVS	1233
QY	711	RRRFVDLFI-----NGGFI-----STRSITVILSEGPVAPNPTIGQD747	
Db	1234	GKRDVGTWCYARGHRAAPADPAFLTGGCLLGDGDTAKAVASRLU---ALGVPTVTGGG	1291
QY	748	APAG-----RTFDGDLARVSVEV---IRDIRVQRVVFSGNCTNLSEAPARLVGLASAYOR	801
Db	1292	RPPGPERYRELVPATRLAVDLWPLFSDASHRGCAAGAVRTAQDAALHLLHLARAFGA	1351
QY	802	QEK3-----VDMLEHGLGFLTKQFHGLLIPPRCMPPNSSKSPNPQWFNTLLQRNQMPADKUT	856
Db	1352	LEERHPARVTVTTGA-----HDVL-----GDDLA1376	

QY	957	HBEITTI-AAVRTEET--AAINFINLPPTICIGELAQFYMANLILKYCDHSQVYLINTLT	913			
Db	1377	HPEHATVFAAAKVIPIREYPMIATCALDVEPLGDAE---RLADLIVRELGAAR---ETTV	1429			
QY	914	SIITGARR---PRDPSSVLHWIKDYTSAADIETQAKALLEKTENIPELWTTAFTSTHL	969			
Db	1430	TACGRRRFTPCP---VQPLFAA---PBRPAVRPGGVYL	1463			
QY	970	VRAANORPMVVLGISISKYHGA---GNNRVFOA-GNWSGLNGKNVCPLFTPDRT	1022			
Db	1464	VCGGLGG---IGHLAEVLGRARTTWLTHRPFPFAPGAWDGLPAGH---PEAAVVR	1515			
QY	1023	RRFTIACPRGCGFICVPGSPSGNRETTLSDO---VGGIIVSGG---	1062			
Db	1516	LRSIAA---TGATVVVRADLTTHDAMRALADEVQAHGFVRCGVHAAAGVPDT	1565			
QY	1063	AMVQLAIYTVRAVAGARQAHQAFDDWLSLTDDEFLARDLEEL	1105			
Db	1566	AGMIQRDRAGTDAALAAKLTGTL-----VLDEVFAHRDLDFL	1603			
RESULT 11						
ABP76682						
ID	ABP76682 standard; Protein; 19938 AA.					
XX	AC ABP76682;					
XX	26-FEB-2003 (first entry)					
DE	Streptomyces viridochromogenes Avi gene cluster polypeptide frame 6.					
XX	Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;					
KW	medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.					
XX	Streptomyces viridochromogenes.					
OS	WC200268436-A1.					
PN	06-SEP-2002.					
XX	24-AUG-2001; 2001WO-EP09815.					
PD	25-FEB-2001; 2001DE-1009166.					
XX	(COMB-) COMBINATURE BIOPHARM AG.					
PR	Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;					
XX	WPI; 2003-018650/01.					
XX	N-PSDB; AB237516.					
XX	New avilamycin derivatives, useful for treatment of infections, and					
PT	nucleic acid encoding avilamycin synthesis enzymes -					
XX	Example 1; Page 68-301; 319pp; German.					
XX	The invention relates to avilamycin derivatives (I) with antibacterial,					
CC	virucide, protozoacide and fungicide activity. (I) are useful for					
CC	treatment of infections (bacterial, viral, protozoal or fungal), in human					
CC	or veterinary medicine, particularly where caused by Staphylococcus					
CC	aureus. (I) are more hydrophilic than known avilamycins. The present					
CC	sequence is that of an avilamycin synthesis enzyme from the Streptomyces					
CC	viridochromogenes Avilamycin A biosynthetic gene cluster					
CC	(AB237515-AB237516).					
XX	Sequence 19938 AA;					
Query Match						
Best Local Similarity 2.0%; Score 125.5; DB 24; Length 19938;						
Matches 246; Conservative 129; Mismatches 451; Indels 445; Gaps 65;						
QY	102	PSSAAPNLTRACNABRPGFSR--CGPPVDGAVETTG-----AETCTRLG-----	146			

CC be used in vaccines for preventing or attenuating an infection caused
 CC by a member of the Enterococcus genus in an animal. They can also be
 CC used for detecting Enterococcus antibodies in a sample. The nucleotide
 CC sequences can be used for detecting Enterococcus nucleic acids.
 CC Products from the present invention can also be used for screening
 CC compounds to identify agonists and antagonists of *E. faecalis* protein
 CC activity.

XX Sequence 1074 AA;

Query Match 2.0%; Score 123; DB 20; Length 1074;
 Best Local Similarity 18.3%; Pred. No. 0.11;
 Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps 52;

QY 250 RVNVTAVARGAHLAPDENHEGAVLPDITVYFQSSSGSTTTTARGARRDNVNSTKPS 309
 DB 67 RTTSIAYENGAKQTVFC-IPGVSIPTEVTHGY-----QKNPL 104

QY 310 PSGGERLASIM--AATALHAEVIFNTGIYEE-----TPTDIKEWPMFI 353
 DB 105 PMSDKAKLVSVLWEKAGTDIDTNVAGKMIWEVGVYKLSIKRLGASVDIK----- 158

QY 354 GWEGTLPRNLALGSYVAR-----VAGVIG-AMVFPNSALYITEVEDSGMTEAKDGGP 405
 DB 159 SIEGKINK--AIEYQKPSFHNTVTKILGQSTTLIDKNELNLSEFDKVVQNTA----- 211

QY 406 GPSNRFYQFAGPHLAANPQTRDRGHVLSQSGTSSNTFSDVYALICGFGAPLLARLL 465
 DB 212 ---NIDRVVIGNQLVLP-----NSNSKSGTLTKKSAGTGP-VAYKK 251

QY 466 FYLERCDAGFTGHHGALKVVTTFSEIPSCIEKHTRPVCAGHTVHLRMRPFQ 525
 DB 252 AGLQTVAGALDKNTVAIKNVETKGS-LKIKIDKESGDIVPTEVPHL-----DFGK 304

QY 526 A-----TROPIGVGMNSQYSCDPLGNVA-----PYLILKPKQDQTEAAKATM 570
 DB 305 ALPSKDVTTDKGI-----SILDGIPGHYKIVITEKSVDPDPNIDTTPMAATIKAGETI 358

QY 571 QDVTYATLERLFDLEORILLDRGAPCSGSLG-----SVIVDHPT---FRILDTLRA 621
 DB 359 SMTSKNRQKQILLKXTG-VETGTDLWNDNYSLAGNTFAIRKDSPAGEIVQBITDEKG 417

QY 622 RIEOTTTFQMKVIVETRDYKIRGLSATHSMALTDP-----YSGAPCPTNPLVKT 675
 DB 418 RAE--TPKELANALELGTYYVTE--TKSSNGFNTFKPTKVELKYANQVALVTSNVKQ 473

QY 676 HLAUVQDLAL-----SQCHCVFYGOV-----EGNFRNQFPVLRFRFVDL 717
 DB 474 NQBITGETTLTKEDKOTGNESQKAEFKGAEYTLFTAKDQAVKWEAFK-----TEL 526

QY 718 FNGGFTSTRITVLSB-GPVSAFNPETLGO-----DAPAGRTFDGLARVVE----- 764
 DB 527 VKGTASDETIVTLALDEKQAVVXHLAINEYFWQETKAPEGVTLDETKYPVSIKKVDNNE 586

QY 765 ---VIRDIRVKNRV-----FSGNCTNLSEAA-----RARLVGLASAYQROK 805
 DB 587 KNAVITRDVTAKEQVIRFGDFDKFAGSADGTAETGFNDLSFKVSPLEGTXTITGAENKA 646

QY 806 VDLHGNALGF-----LKQPHGLLFFPGM-----PPNSKS-----PNP 838
 DB 647 TTACNEQLGQDGYGKFPENLPYGYLLLEIEA---PEGFKITPLEIRSTRFFKKNDDYAKS 703

QY 839 QWFTWLLQRNQ-----MPADKLTHEITITIAVKRETEYAAINFNLN-PTCIGE 889
 DB 704 EYVFTITEBQKQPIKMWTVPEKLNNE-----FVSLNRLMLYDLPEKEDSITS 754

QY 889 LAQFYVANLILKVDHSHQVILNLTSTIITGARRRPPSVLHWIRKD---VTSAAIDITQ 945
 DB 755 LATWKGNGKLANLDFTE-LVDKL-----RYNLHEIKEDWYVAQAIDVEA- 799

QY 946 AKALLEKTENLPELWTTATFTST-----HLVPAAMNQPMVVLGISISKYHGA 992
 DB 800 TKAAQEKDEKAPVVAETATLANKEKGTWKLHLKLTAEQ-----VLDKSIVLFNIV 853

QY 993 AGNRVFOAGNWSGLNGKKNVCPLFTEDTRRFTIACPRGGFICVTPGSSGNRETTILSD 1052
 DB 854 YENKVAFEAGNE-----PVA-----KDAISLNN 875

QY 1053 QVRGIIVSGGAMVQLAIYATVVRVARGAR-----AQHMAFPDDWLSTDDDEFIARDL-- 1102
 DB 876 QAQ-----TVNCTIERHVSIOQKAHLEDGSGTFTTHGDMDFDVSVTHDVL 923

QY 1103 --EELHDIQIQTLETPWTVEGALBAVKILDE-----KTTAGD-----GETPTNLAFNFD 1149
 DB 924 GSKEAFETILYALLPDGNTKEIKWSGKIEHEVNMKEFTKTVLAEKVDTGKYPEGTKFTF- 982

QY 1150 SCEPSHDTISNV-----LNISSNISGSTVPGIKRPPEDDE 1185
 DB 983 -TEINYEKDGNGVGNKHEDLKEKSQTLTPKEVPIPTPKQPE 1024

RESULT 14

AA000188
 ID AAY00188 standard; Protein; 1074 AA.
 XX AAY00188;
 XX 20-APR-1999 (first entry)
 XX Enterococcus faecalis protein EF094.
 DE Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
 KW detection; attenuation; antigenic.
 XX Enterococcus faecalis.
 XX WO950554-A2.
 PD 12-NOV-1998.
 XX 04-MAY-1998; 98WO-US08959.
 PR 14-NOV-1997; 97US-0066009.
 PR 06-MAY-1997; 97US-0044031.
 PR 16-MAY-1997; 97US-0046655.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Bailey C, Choi GH, Hromockyj A, Kunsch CA;
 DR WPI; 1999-070095/06.
 DR N-PSDB; AAX20178.
 XX New isolated Enterococcus faecalis polynucleotides - used to develop
 PT products for the detection of Enterococcus and for use in vaccines
 PT for prevention or attenuation of Enterococcus infection
 XX Claim 9; Page 186-187; 30pp; English.

The present sequence represents a protein isolated from
 Enterococcus faecalis. The present invention describes genes, proteins
 and antigenic polypeptides isolated from *E. faecalis*. The proteins can
 be used in vaccines for preventing or attenuating an infection caused
 by a member of the Enterococcus genus in an animal. They can also be
 used for detecting Enterococcus antibodies in a sample. The nucleotide
 sequences can be used for detecting Enterococcus nucleic acids.
 CC Products from the present invention can also be used for screening
 CC compounds to identify agonists and antagonists of *E. faecalis* protein
 CC activity.

XX Sequence 1074 AA;

Query Match 2.0%; Score 123; DB 20; Length 1074;
 Best Local Similarity 18.3%; Pred. No. 0.11;
 Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps 52;

QY 250 RVNVTAVARGAAHLAFDENHEGAVLPDPITYTTFQSSSGTTTARGARRNDVNSTKPS 309
Db 67 RTTSLVAEYNGAKQTVFC-IEPGVSIPTVTHGY-----QKNPL 104
QY 310 PSGGFERRLASIM--AADTALHAENVFNTGIYEE-----TPTDIKWPMFI 353
Db 105 PMSDKAKLVSVLWEKAGTDIDTNNVAQKMIWEVNGYKLSIKRLGGASVDIK-----158
2Y 354 GMEGTLPRLNALGYSYAR-----VAGVIG-AMVFPNSALYLTVEVDSGMTAKDGGP 405
Db 159 STEGKINK--AIEEYQKPSFNTTIVKTLGQSTTLIDKNELNSEFDKVVQNTA-----211
2Y 406 GPSFNRFYQFAGPHLAANPQTRDCHVLSQSTGSSNTFSDVYALICGFGAPLLARLL 465
Db 212 ----NIDYRVIGNQLVLT-----NSNSKSGTTLKKSAGTGTP-VAYKK 251
2Y 466 FYLERCDAGFTGGHDALKYVGTGPDSEIPCSLCEKHTRPVCAHTVHRLRQRPFRGQ 525
Db 252 AGLQTVWAGALDKPNTYAIKINVETKGS-LKIKIDKESGDIVPETVFLH-----DFGK 304
2Y 526 A-----TROPIGVFGTMMNSQYSDCDPLGNVA-----PYLILRKPQDQTEAAKATM 570
Db 305 ALPSKDVTTDKGI-----SILDGIPHGTKVITEKSVDPDYMDITTPMAATIKAGETI 358
2Y 571 QDTYRATLERLFDLEQEBLLDRGAPCSSEGLS-----SVIVDHT-----FRRILDTLRA 621
Db 359 SMTSKNMROKQGLLEKTG-VETGTLMDNDYSLAGNTFAIRKDSPAGEIVQEIITDEKG 417
2Y 622 RLEQTTQFMKVLVETRDYKIREGLSEATHSMALTFDP-----YSGAFCPITFLVKRT 675
b 418 RAE--TPRELANALELGTYYVTE--TKSNGFVNTFKTKVELKYANGTVALVTSNVKQ 473
Y 676 HLAUVQDLAL-----SOCHCVFYGOV-----EGRNFRNQFQVLRARRFVDL 717
b 474 NOBITGETTLTKEDKDTGNESQKAEFGKAEYTLFTAXDGOAVKWEAFK-----TEL 526
Y 718 FNGGFISTRSIVTLSE-GPVSAPNPTLGO-----DAPAGTFDGLARVSVE-----764
b 527 VKGTKASDETIVLADENQNAVHKLAINFVWOSTKAPBGTTLDETYPVSIKKVDNNE 586
Y 765 ----VIRDIRVKNRVV-----FSGNCTNLSEAA-----BARLVGLASAYQREKR 805
b 587 KNAVITRDVTAKEQVIRGDFDFKAGSADGTAEFGNDLSFKVSPLEGTKEITGADKA 646
Y 806 VDMHLGALGF-----LLKQFHLLPFRGM-----PNPSKS-----PNP 838
b 647 TTACNEQLGFDGKGFENLPYGVDLLEBIEA---PEGFQKITPLEIRSTRFENKDDYAKS 703
Y 839 QWFMTLLQRNQ-----MPADKLTHEEITIAAVKRFTSEYAAINFILP--PTCIGE 888
b 704 EYVFTITEEGQKQPKVMVTVYEXLITNE-----FVSLNRLMYDLPEKEDSLTS 754
Y 889 LAQFYMANLILKYDHSQYLINTLTSITGARRPRDPSSVLHWIRKD---VTGAADIETQ 945
b 755 LATWQGNKGLNTLDFTE-LVDKL-----RYNLHEIKEDWYVVAQAIDVEA- 799
Y 946 AKALLEKTENLPWLTTAFTST-----HLVRAAMQRPWVLGISIKVHGA 992
b 800 TKAAQEKDEKAPVVAETATLANKEKTGTWKILHLKLTAEQ-----VLDKGIIVFNYY 853
Y 993 AGNRRVFOAGNWSGLNGKNVCPFLTPDRTRRFIACPRGGFCIPVTPGSSGNRETTLS 1052
b 854 YENKVAPEAGNE-----PVA-----KDALSLNN 875
Y 1053 QVRGIIVSGGAWQVLAIVATVVRVARGAR-----AQNMAFDWLSTLDEFLARDL--1102
b 876 QAQ-----TVNCTIERHVSQTKAHLEDGSGTFTHGVDVMDFDVSVTHDVL 923
Y 1103 --EELHQI11QTLFPTWTCALAEVAKILDE-----KITAGD-----GTPTNLAFND 1149
b 924 GSKEAFETILVALPDCGTNKEINWKSQKIEHVNDKSEFTKVLAEKVDYTKYPGTGTFP- 982
Y 1150 SCEPSHDTTSNV-----LNISGNSISGTVFGLKRPPEDDE 1185

Db 993 -TEINYEKNGVNGKHNEDLKESQTLTPKEVPTIPSTPKQPE 1024

RESULT 15

ABP43407

ID ABP43407 standard; Protein; 1074 AA.

XX ABP43407;

XX 05-AUG-2002 (first entry)

XX E faecalis EF094 protein.

XX Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic.

XX Enterococcus faecalis.

XX US2002045737-A1.

XX 18-APR-2002.

XX 04-MAY-1998; 98US-0071035.

XX 04-MAY-1998; 98US-0071035.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Choi GH, Bailey C, Hromockyj A, Kunsch CA;

XX WPI; 2002-425450/45.

XX N-PSDB; ABN98163.

XX New genes and polypeptides from Enterococcus faecalis, useful as
XX vaccines for preventing, treating or attenuating an infection caused by
XX a member of the Enterococcus genus in an animal, particularly E.
XX faecalis

PS Claim 9; Page 163; 255pp; English.

XX The present invention provides the protein and coding sequences of a
XX number of polypeptides from Enterococcus faecalis. The proteins can be
XX used as vaccines for preventing or attenuating an infection caused by a
XX member of the Enterococcus genus in an animal, particularly E. faecalis.
XX The polynucleotide is also useful for preventing or treating E. faecalis
XX infection. The present sequence is a protein of the invention.

SQ Sequence 1074 AA;

Query Match 2.0%; Score 123; DB 23; Length 1074;
Best Local Similarity 18.3%; Pred. No. 0.11;
Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps 52;

QY 250 RVNVTAVARGAAHLAFDENHEGAVLPDPITYTTFQSSSGTTTARGARRNDVNSTKPS 309
Db 67 RTTSLVAEYNGAKQTVFC-IEPGVSIPTVTHGY-----QKNPL 104
QY 310 PSGGFERRLASIM--AADTALHAENVFNTGIYEE-----TPTDIKWPMFI 353
Db 105 PMSDKAKLVSVLWEKAGTDIDTNNVAQKMIWEVNGYKLSIKRLGGASVDIK-----158
QY 354 GMEGTLPRLNALGYSYAR-----VAGVIG-AMVFPNSALYLTVEVDSGMTAKDGGP 405
Db 159 STEGKINK--AIEEYQKPSFNTTIVKTLGQSTTLIDKNELNSEFDKVVQNTA-----211
QY 406 GPSFNRFYQFAGPHLAANPQTRDCHVLSQSTGSSNTFSDVYALICGFGAPLLARLL 465
Db 212 ----NIDYRVIGNQLVLT-----NSNSKSGTTLKKSAGTGTP-VAYKK 251
QY 466 FYLERCDAGFTGGHDALKYVGTGPDSEIPCSLCEKHTRPVCAHTVHRLRQRPFRGQ 525
Db 252 AGLQTVWAGALDKPNTYAIKINVETKGS-LKIKIDKESGDIVPETVFLH-----DFGK 304

QY 526 A-----TRQPIGVFGTMSQYSDCDPLGNYA-----PYLILRKGDOQTEAKATM 570
 Db 305 ALPSKDVTTDKOGI-----SILDGPHGKTKVITEKSVDPDMIDTTPMAATIKAGETI 358
 QY 571 QTYRATLERLFDLEQERLLDRGAPCSSEGLS-----SVIVDHT---FRRILDTLRA 621
 Db 359 SMTSKNMROKQOILLEKTC-VETGTDLMNDNYSLAGNTFAIRKDSPAGEIVQEITTDK 417
 QY 622 RIEQTTQPMKVLVTRDYKIREGLSEATHSWALTFDP-----YSGAPCPIITNLFVKRT 675
 Db 418 RAE--TPRELALAELELYVTE--TKSNGFVNTFKTKVELKYANQTVALTNSVKQG 473
 QY 676 HIAVQDLAL-----SOCHCVFYGOQV-----EGNPRNQFQVLRERRFVDL 717
 Db 474 NQEIITGETTLTKEDKDTGNESQKAEFKGAETLFTAKDQAVKWSFAK-----TEL 526
 QY 718 FNGGFSTRSIIVTISE-CPISAPNPTLGO-----DAPAGRTFGDLARVSE----- 764
 Db 527 VNGTKASBETVTLADEKNQVAVKHLAINEFWQETKAPEGYTLDETYPVSIKKVONNE 586
 QY 765 ----VIRDIRVKNRVV-----PSGNCNTNLSEAA-----RARLVGLASAYQROEKR 805
 Db 587 KNAVITRDVTAKEQVIRFGDFKFPAGSADGTAEFGNDLSFKVSPLEGYXETGAEDKA 646
 QY 806 VMLHGLGFP-----LLKQFHGLLPPRG-----PNSKS-----PNP 838
 Db 647 TTACNEQGLGFGYKFNLPYGYLLEIEA---PEGFKITPLEIRSTFKENKDDYAKS 703
 QY 839 QWFTLLORNO-----MFADKLTHEIITIAAVKFTBEVYAAINFILP--PTCIGE 888
 Db 704 EYVFTITEGQKQPKVTVVEKLTNNE-----FVSLNRLMLYDLPEKBSLTS 754
 QY 889 LAQFYMANLILKYODHSQYLNTLTSITGARRPRDPSSVLHWIRKD---VTSAAADIETQ 945
 Db 755 LATWKGNGKKLNTLDFTE-LVDKL-----RYNLHEIKEDWYVVAQAI DVEA- 799
 QY 946 AKALLEKTENLPWLTTAFTST-----HLVRAAMQORPMVVLGISIKYHGA 992
 Db 800 TKAAQEKDEKAKPVVIAETTATLANKEKTGTWKLHLKTAEO-----VLDKSI VLFNYV 853
 QY 993 AGNNRVFOAGNWSGLNGKKNVCPLETFDRTRRFIACPRGGFCPTGPPSSGNRETTLS 1052
 Db 854 YENKVAFEAGNE-----PVA-----KDASLNN 875
 QY 1053 QVRGIIYVSGAMWOLAIYATVRAVGAR-----AQHVAFDDMLSLTDDDEFARDL-- 1102
 Db 876 QAC-----TVNCTIERHVSITQKAHLEDGSGTTFHGDVMDMFDVSVTHDVL 923
 QY 1103 --EELHQIITQLETPWTVGALEAVKILDE-----KITAGD---GETPTNLAFNED 1149
 Db 924 GSKEAFETILYALLPDGTNKEIWKSGKIEHVNDKBEFTKTVLAEKVDTKYPEGTKFTF- 982
 QY 1150 SCFSPSHDTTNSV-----LNISGSNISGSTVPGLKRPPEDDE 1185
 Db 983 -TEINVEKDNVAGKHNEDLKERSQTLTPKEVPTIPSTPKQPE 1024

Search completed: January 30, 2004, 13:12:34
 Job time : 52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

MM protein - protein search, using sw model

run on: January 30, 2004, 13:14:08 ; Search time 44 Seconds

(without alignments)

5682.099 Million cell updates/sec

itle: US-09-769-699-2

effect score: 6294

sequence: 1 MENTQKTVTVTGLGVVYA.....DELFDLGIPIKGNITMEM 1203

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 789580 seqs, 207824079 residues

total number of hits satisfying chosen parameters: 789580

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : Published Applications AA:

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pap.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pap.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pap.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pap.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pap.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pap.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pap.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pap.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pap.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pap.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pap.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pap.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Match	Length	DB ID	Description
1	3054	48.5	1452	14	US-10-050-673-2
2	3015.5	47.9	1196	12	US-10-200-562-200
3	3015.5	47.9	1196	12	US-10-237-551-200
4	3015.5	47.9	1196	12	US-10-237-551-232
5	788	12.5	274	12	US-10-237-551-231
6	736	11.7	248	12	US-10-237-551-230
7	123.5	2.0	850	15	US-10-456-761-9121
8	120.5	1.9	2597	10	US-09-305-129-2
9	120.5	1.9	2597	10	US-09-305-129-10
10	120.5	1.9	2597	10	US-09-305-129-13
11	120.5	1.9	2597	10	US-09-305-129-13
12	120.5	1.9	2597	10	US-09-305-129-13
13	120.5	1.9	2597	10	US-09-305-129-13
14	120.5	1.9	2597	10	US-09-305-129-13
15	118	1.9	19608	12	US-10-084-846A-8

16	114.5	1.8	1496	12	US-10-021-660-125	Sequence 125, App
17	114.5	1.8	1496	12	US-10-331-456A-28	Sequence 28, Appl
18	113.5	1.8	1477	12	US-10-274-583-20	Sequence 20, Appl
19	112	1.8	6620	12	US-10-080-334-290	Sequence 290, App
20	111	1.8	5215	10	US-09-861-289-2	Sequence 2, Appl
21	111	1.8	5215	10	US-09-860-846-2	Sequence 2, Appl
22	111	1.8	5215	11	US-09-988-384B-2	Sequence 2, Appl
23	111	1.8	5215	11	US-09-836-821-2	Sequence 2, Appl
24	111	1.8	5215	12	US-10-271-889-45	Sequence 45, Appl
25	110.5	1.8	3069	10	US-09-712-363-246	Sequence 246, App
26	109.5	1.7	2630	14	US-10-077-130-2	Sequence 2, Appl
27	109.5	1.7	7968	14	US-10-077-130-5	Sequence 5, Appl
28	108	1.7	499	12	US-10-369-493-10577	Sequence 10577, A
29	108	1.7	2596	15	US-10-307-019-6	Sequence 6, Appl
30	107.5	1.7	8026	12	US-10-132-134-12	Sequence 12, Appl
31	107	1.7	505	12	US-10-369-493-2550	Sequence 2550, App
32	106.5	1.7	1299	15	US-10-156-761-12449	Sequence 12449, A
33	106.5	1.7	3298	12	US-10-174-677-8	Sequence 8, Appl
34	106.5	1.7	3298	12	US-10-120-801-51	Sequence 51, Appl
35	106.5	1.7	3298	15	US-10-160-758-16	Sequence 16, Appl
36	105.5	1.7	707	15	US-10-156-761-9979	Sequence 9979, App
37	105.5	1.7	1563	12	US-10-334-443-34	Sequence 34, Appl
38	105.5	1.7	1967	15	US-10-219-834-85	Sequence 85, Appl
39	105.5	1.7	1967	15	US-10-225-567A-575	Sequence 575, App
40	104.5	1.7	635	11	US-09-851-847-5	Sequence 5, Appl
41	104.5	1.7	1471	7	US-08-811-519A-1	Sequence 1, Appl
42	104.5	1.7	1515	12	US-10-240-154-8	Sequence 8, Appl
43	104.5	1.7	6304	15	US-10-147-026-16	Sequence 16, Appl
44	104	1.7	3192	12	US-10-132-134-10	Sequence 10, Appl
45	103	1.6	3472	12	US-10-029-120-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-10-050-673-2
; Sequence 2, Application US/10050673
; Publication NO. US20020151033A1
; GENERAL INFORMATION:
; APPLICANT: David M. Knipe
; APPLICANT: Travis J. Taylor
; APPLICANT: Elizabeth McNamee
; TITLE OF INVENTION: Replication-Competent Virus Expressing A
; FILE REFERENCE: Fusion Protein
; FILE REFERENCE: HU98-05
; CURRENT APPLICATION NUMBER: US/10/050.673
; CURRENT FILING DATE: 2002-01-16
; PRIOR APPLICATION NUMBER: US/09/127.227
; PRIOR FILING DATE: 1998-07-31
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FASTSQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1452
; TYPE: PRT
; ORGANISM: herpesvirus
US-10-050-673-2

Query Match 48.5%; Score 3054; DB 14; Length 1452;
Best Local Similarity 49.8%; Pred. No. 8.3e-294;
Matches 609; Conservative 200; Mismatches 366; Indels 48; Gaps 17;

QY	1	MENTQKTVT---VPTGPIGVYV--ACRVEDLDLEISFLAARSTDSIALPLMRNLTVE 55
DB	1	METKPKTATTTKVPPIGVYVACPCSEGL--LALLSARSGADAVAPLVGLTVE 58
QY	56	KTFSSLLAVSGARTTGAGAGITLKTTHSFYFVSFVHGKHLPSSAPNLTACNA 115
DB	59	SGFEANVAVVGSRTTGLGTAVSKLUTPSYSSVYVFGHRLDSTQAPNLTUCER 118
QY	116	AREPFGFRCGCPVDCAVETTGAEITRIGLEPENILVIVTALFKFAVFCNVFLHY 175
DB	119	ARRHFGFSDYTPRGDLKHETTGAALCERGLDPRDALLVIVWTEGKFAVCINNTEHL 178

09/769,699
Search Notes

176 GGLDIVHNGDVIRIPLFPVQLFMPDVRNLVLPDFFNTHRSIGEGFVYPTTFYNTGLCH 235
179 GGSOKVTIGGAHVIRIPYVQLFMPDFFSRVIAFPFNANHRISGENFTYPLPFNRFLNR 238
236 LIHDCVIAPMAVALVRNVAVAGAAHLAFDENHGAVALPPDITYTYFOSSSGTATTAR 235
239 LLFEAVGPAVALRCRNDVAVARAHLAFDENHGAVALPADITTFATFASQ--KTPR 236
296 GARNDVNSTSKPSSGFERLASIMAAUTALHAETVNTGIEETPTDIKEWPMFIGM 355
297 GGR-----DGGKGPGAGFEORLASVAGDAALALESVMAVDEPTDISAWPLCEGQ 351
356 EGTLPRLNALGSYARVAGVIGAMVFPNSALYTEVEDSGMTEADGGPGPFRFYQF 415
352 DTAARANAVAGYLAAGLVGAMVFTNSALHTEVDDAGPADPKDHSK-PGFYFFLV 410
416 AGPHLAANPQDRDGHVL-----SSQSGSSNTEFSDYLALICGFGAPILARLLFYL 468
411 PGTHVAANPOVDREGHVVPFEGRPTAPLVGGTQ-EFAGEHLANLCOFSPALLAKWLFYL 469
469 EBCDAGFTGGH-G-DALKYVGTGFDSEIPCLCEKHTRPVCAHTVHRLRQMRPRGOAT 527
470 ERCDGGVIVGRQEMDVRYVADSNQTDVPCNLCTFDRHACVHTTLMRLRARHPKASAA 529
528 RQPIGVFTWNSQVSDCDPLGNVAPYLILKPGDQTEAAKATMODTYRATLERLFDLEQ 587
530 RGAIGVFTWNSMVSQVSDCDVLGNYAASALKR-ADGSETARTINQETIYRATERVVALEET 588
588 ERLLDRCAPSSGSLSVIIVDHTFTFRILDLRARIEQTTQFMKVLVTRDYKIREGLS 647
589 LQYVDQAVPTAMGRLETTIITREALHTVNNVQVVDREVEQLMRLNVEGRNFKPRDGLG 648
648 EATHSMALTDPDPSGAFCPITNELVKRTHLAVVODLALSOCHCVFQCOQVEGRNRFNOFO 707
649 EANHMSLTLDPAACGCPILQLGRSLAVVQDLALSQCHGVFAGQSGVEGRNRFNOFO 708
708 PVLRRPVDLFGNGFISTRSITVTLSEG-PVSAFNPTLQGDAPAGRTFDCDLARVSVEVI 766
709 PVLRRRVMDMNFNGFTLSAKTLTVALSEGAACAPSUTAGTAPAESFSGDVARVTLGFP 768
767 RDLRVNRVFNESCNTLSEBARLVGLASVQROKREYDMLHGAFLKQFGLLFP 826
769 KEURVSRVLFGASANASAAKARVASQSVQKEDKEDVILLGLFLKQFHAAIFP 828
827 RGMPNPKSPNPQWFTLLQRNQMPADKLTHEBITTAAVKPTEBYEYALNFINLPPTCI 886
829 NGKPPGSGNPQWFTALQRLPOLPARLSREDIETIAFKFSLDYGAIFNINLAPNV 888
887 GELAQFYMANLILKYCDHSOYLINTUTSIITGARRPRDPSVLHWRKDVTSAADTQA 946
889 SELAMTYMANQLIRYCDHSTYFINTUTAIAGSRPPSVQAAAAW---SAQCGAGLEAGA 945
947 KALLEXTENLPELWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNRVFOAGNWSG 1006
946 RALMDAVDAHGPAGWNTSFPASCNLLRPVMAARPMVVLGISISKYHGAAGNRVFOAGNWS 1005
1007 LMGKKNVCPLFTDRTRFTIACPRGGFCPTVGPSSGNRETTLDQVRGIIVSGAMVQ 1066
1006 LMGKKNACPLIFDRTRKFKVACPRAGFYCAASNLCGGGAHSSLCQLRGIISEGGAVA 1065
1067 LAITYAVRAVAGARQAHMAFDWLSITDDEFARDLREHLHQIOTLTPWYVEGAL--- 1123
1066 SVFVATVKSGLPRTQOQLQIEDWLLADEYLSUEEMELTARALGRNGRWSVTAALAVA 1125
1124 -BAVKILDEKTTAGDGETPTNLAFNFD--SCBPSHDTTNSVINTSGSNISSGTVPLKRP 1180
1126 HEAEALVSQLNAGE-----VFNFGDGC-----DDNATPGGPGAPGAPAGKGA 1173
1181 PEDDELFDLSGIPKXGNITMEM 1203
1174 FHGDOPFG-EGPPDCKGDJTLDM 1195

RESULT 2
US-10-200-562-200
; Sequence 200, Application US/10200562
; Publication NO. US20030165819A1
; GENERAL INFORMATION:
; APPLICANT: McGowen, Patrick
; APPLICANT: Hosken, Nancy A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C2
; CURRENT APPLICATION NUMBER: US/10/200.562
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 200
; LENGTH: 1196
; TYPE: PRT
; ORGANISM: HSV2
US-10-200-562-200

Query Match 47.9%; Score 3015.5; DB 12; Length 1196;
Best Local Similarity 49.0%; Pred No. 4e-290;
Matches 595; Conservative 213; Mismatches 366; Indels 41; Gaps 16;

QY 4 TOKTVTPTGTPGLGVY--ACRVEDLDLEISFLAARSTDSDLALPLMRLNLTVEKFTTS 61
DB 7 TTTTVKVPPOGPMGVYGRACPAEGLEL--LSLLSARSGDADVAVALVGLTVESGFEAN 64
QY 62 LAVVSGARTTGLAGAGITLKTTSHFVPSVFEHGGKHVLPSSAANLACRACNAARERG 121
DB 65 VAAVVGSTTGLGTAVSLKLMHSVSPSVYFHGGHLLAPSTQANLRLCERARPHFG 124
QY 122 FSRQGGPPDVGAVTTGAEICTRLGLEPENTILYLVVTALFKEAVMCNVFLHYGLDIV 181
DB 125 FADYAPRCDLKHEITGDALCEGLDDEPRALYLVITEGFEAVCISNTEHLGMDKV 184
QY 182 HINHGDIVRILPLFPVQLFMPDVRNLVDPDNTHRSIGEGFVYPTTFYNTGLCHLHDCV 241
DB 185 TIGDAEVRHRIPTVPLQFMWDFSRVIADPFCNHRISGENFNYPFPFNRPLARLFEAV 244
QY 242 IAFMAVALRVNTAVARGAAHLAFDENHGAVALPPDITYTYFOSSSGTITARGARRND 301
DB 245 VGPAVALRARNVDAAVAAAHLAFDENHGAVALPADITTFATFASQ--KPGGAR--- 299
QY 302 VNSTKPSGSGFERLASIMAAUTALHAETVNTGIEETPTDIKEWPMFIGMCTLPR 361
DB 300 --DAGNKGPGAGFEORLASVAGDAALALESVMAVDEPTDITWPLLEGQETPAAR 357
QY 362 LNALGSYTVARVAGVIGAMVFPNSALYTEVEDSGMTEADGGPGSPFNRFYQFAGPILA 421
DB 358 AGAVGAYLARAAGLVGAMVFTNSALHTEVDDAGPADPKDHSK-PSFYRFTLVPGTHVA 416
QY 422 ANPQTRDQGHVL-----SSQSGSSNTEFSDYLALICGFGAPILARLLFYLERCDA 474
DB 417 ANPQDREGHVVPFEGRPTAPLVGGTQ-EFAGEHLANLCOFSPALLAKWLFYLERCDGG 475
QY 475 AFTGGHG-DALKYVGTGFDSEIPCLCEKHTRPVCAHTVHRLRQMRPRFGAIFQPIGV 533
DB 476 VIVGREGMDVFRVADSGDQVPCNLCTFETKHAHTTLMRLRARHPKFAAARGATGV 535
QY 534 FGPMNSQVSDCDPLGNVAPYLILKPGDQTEAAKATMODTYRATLERLFDLEORLDR 593
DB 536 FGPMNSAYSDCDVLGNYAASALKR-ADGSENTRTIMOSTYRAATERVMAELEALQYVDQ 594
QY 594 GAPSSEGLSSVIVDHTFTFRILDLRARIEQTTQFMKVLVTRDYKIREGLSEATHSM 653
DB 595 AVPTALGRLETIGREALHTVNNIKQVREVBQMLNTEGNEFKFRDGLAEANHAM 654
QY 654 ALTFDPSYGAFCPIITNFLVKRTHLAVVODLALSOCHCVFQCOQVEGRNRFNOFVLRR 713
DB 655 SLSDPYTCGCPCLQLLARRSNLAVYQDLALSQCHGVFAGQSGVEGRNRFNOFQVPLRR 714

QY	714	FVDLENGGPISTRSITVTLSSEG-FVSAPNPTLQDAPAGTTFDGLARVSVVEVIRVK	772
Db	715	VMDLFNNGFLSAKTLUTVALSAGAAICAPSLTAGTAPAESFEGDVARVTLGFKELRVK	774
QY	773	NRVFSGNCNTLSEAAARLVGLASAYQROEKRVDMHLHGALGFLKQFHGLLFPFGPPN	832
Db	775	SRVLFAGASANAASEAAKARVASLOSAYQKPKRVDLLGLPLGFKQFHAVIFPNGKPPG	834
QY	833	SKSPNQFWTLLOQNPADKLTHEITIAAVKRTBYAIAINFLPPTCIGELAQF	892
Db	835	SNQPNQFWTLQORNOPLARLSREDIETIAFKRFSLDYGAINFLAPNNVSELAMY	894
QY	893	YMANLILKYCDHSQYLINTLSITTGARRPRDSSVLHWKDVTSAADTETOKALLEK	952
Db	895	YMANQILAYCDHSTYFINTLTAVIAGSRPPSVQAAAAAPQ---GGAGLEAGARALMDS	951
QY	953	TENLPMLWTATFTHLVRAAMNORPMVILGISIKYHGAAGNNRVFOQAGNSGLNGGN	1012
Db	952	LDHFGAWTSMFASCNLLRPVMAARPMVILGLSISKYMGAGNDRVFOAGNWSLGGKN	1011
QY	1013	VCPLTFDTRRFTIACPRGGFCPTVTPSSGNRETTLSQVKGIIIVSGAMVQLAIYAT	1072
Db	1012	ACEPLIFDTRKFKVLACPRAGFVCAASSLGGGAHSLCEQLRGLIAGGAAVASSVFWA	1071
QY	1073	VYRAGARACHMAEDDWLSITDDDEFIARDLBEHDOIIQTLETPWTVEGAL----	1128
Db	1072	TVKSLGRTQQLQIEDWALLEDEYISEEMESTTRALERHGEWSTDAALEVAHEAEAL	1131
QY	1129	LDEKTTAGDGETPTNLAFNFDSCPSHDTTNSVLNLSGNSISGSTVGLKRPPEDELFD	1188
Db	1132	VSQLAGAA-----VFNFGDGEDDDHAASFGGLAAA--AGAAGVARKRAFHGDDPFG	1182
QY	1189	LSGPIPKHGNITMEM	1203
Db	1183	EGPEKK-DITLDM	1195

RESULT 5

US-10-237-551-231

; Sequence 231, Application US/10237551

; Publication No. US20030165820A1

; GENERAL INFORMATION:

; APPLICANT: Day, Craig H.

; APPLICANT: Hosken, Nancy A.

; APPLICANT: Parsons, Joseph M.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND

; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION

; FILE REFERENCE: 210121.538C3

; CURRENT APPLICATION NUMBER: US/10/237,551

; CURRENT FILING DATE: 2002-09-06

; NUMBER OF SEQ ID NOS: 254

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 231

; LENGTH: 274

; TYPE: PRT

; ORGANISM: Herpes simplex virus

US-10-237-551-231

Query Match 12.5%; Score 788; DB 12; Length 274;

Best Local Similarity 56.5%; Pred. No. 2.6e-69;

Matches 153; Conservative 41; Mismatches 75; Indels 2; Gaps 1;

QY	28	LEBISFLAARSTDSLLALPLMRNLTVKTKTSSIAVVSAGATTGLGAGATLKLTTSHF	87
Db	2	LELLSLSARSADAVAVAPLIVGLTVESGFANVAVGSRITGLGTAVSLKMFPSHY	61
QY	88	YPSVVFHGGKVLPSAAPNLTACNAARERFGSRQCGPPVDGAVETTGAEICTRLGL	147
Db	62	SPSVYVFHGGHLSAPSTQAPNLTCLCERARHFGPSDYAPRCDLKHETTGDALCERLGL	121
QY	148	EPENTLILVWTALFKAEVFCMVFHYGGDLIVHINHGDVIRIPLFPVOLFPDVRNLV	207

US-10-237-551-232

; Sequence 232, Application US/10237551

; Publication No. US20030165820A1

; GENERAL INFORMATION:

; APPLICANT: Day, Craig H.

; APPLICANT: Hosken, Nancy A.

; APPLICANT: Parsons, Joseph M.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND

; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION

; FILE REFERENCE: 210121.538C3

; CURRENT APPLICATION NUMBER: US/10/237,551

; CURRENT FILING DATE: 2002-09-06

; NUMBER OF SEQ ID NOS: 254

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 232

; LENGTH: 1196

; TYPE: PRT

; ORGANISM: Herpes simplex virus

US-10-237-551-232

Query Match 47.9%; Score 3015.5; DB 12; Length 1196;

Best Local Similarity 49.0%; Pred. No. 4e-290;

Matches 595; Conservative 213; Mismatches 366; Indels 41; Gaps 16;

QY	4	TQKTVTPGPGYVY--ACRVEDLDLBEISFLAARSTDSLLALPLMRNLTVKTKTSS	61
Db	7	TTTTVKVPPGPMGYVGRACPAEGLEL--LSLSARSADAVAVAPLIVGLTVESGFAN	64
QY	62	LAVVSGARTGLAGAGITLKTTHSFYSVVVFHGGKVLPSAAPNLTACNAARERFG	121
Db	65	VAAVGSRTTGLGGTAVSLKMFPSYSVVVFHGGHLSAPSTQAPNLTCLCERARHFG	124
QY	122	FRCQGPVVDGAVETTGAEICTRLGLEPENTLILVWTALFKAEVFCMVFHYGGDLIV	181
Db	125	FADYAPRCPDLAKHETGDCALCERLGLDPRALLYLIVITEGFEAVCISNTFLHGGNDKV	184
QY	182	HINHGDVIRIPLFPVOLFPDVRNLVPPFNTHRSIGEGFVYPTPTPYNTGLCHLHDCV	241
Db	185	TIGDAEVHRIPIVPLQMFOPFSRVIAOPFNCHRSIGENFNPLPFNFRPLARLLFEAV	244
QY	242	TAPMAVALRVNVTAVARGAAHLAFDENHEGAVLPDITVITYFOSSSGTTARGARRND	301
Db	245	VGPAVALRARNVDVARAAAHAFDENHEGAALPADITTFEASOG--KPGGAR---	299
QY	302	VNSTSKPSPSGFERRLASINAAATLHAEIFNTGVIETPTDIKEWFIOWEGTLPR	361
Db	300	--DAGNKPAGFQORLASVWAGAAALASISVMAVDEPDPDITWPLLEQETFAAR	357
QY	362	LNALGSYIARVAGVIGAMVFSNALVTEVEDSGMTAEKDGPGPSFNRFYQFAGPHLA	421
Db	358	AGAVGAYLARAAGLVAMVFSTNSALHLTEVDVDDAGPADPKDHSK-PSFYRFLVPQTHVA	416
QY	422	ANPOTRDRGHVL-----SSQSTGSSNTFESVDYLALICGFGAPLALLFVLERCDAG	474
Db	417	ANPQDRGHRVVGEGRPTAPLVGGTQ-EPFAGHSLAMLCGFSFALLAKMLFYLERCDGG	475
QY	475	AFTGGHG-DALKYVTGTFDPSBIPCLCEKTRPVCAHTTVHRQRMPRFGQATRPQIV	533
Db	476	VIVGQEMDVFRYVADSQDTPVNCNLCTFETRHACATTLMLRAKHPFASARGAIV	535
QY	534	FGTWNQSYSCDPLGNVAPYLILKPGQDTEAKATVQDTYRATLERLFDILEQERLLDR	593
Db	536	FGTWNAYSQCDVLGNVAAFSALKR-ADGSENTRTIMQETYRATLERNVMAELEALQYVDQ	594
QY	594	GAPCSSEGLSSVVDHPTFRILDLTAAIREQTTTQFMKVLVETRDYKIREGLSATHSM	653
Db	595	AVPTALGRLETTIGNREALHTVNNIQLVDREVEQLMRNLIEGRNFKFRDGLAENHAM	654
QY	654	ALTDPYSGAFCPITNPLVXETHLAVVDLALSOCHCVFVQGOVEGRFNPNFQVPLRRR	713
Db	655	SLSDPTCGCPCLQLLARNSLVQDLALSOCHGVFAGOSVEGRFNPNFQVPLRRR	714

Db 122 DDPRLLYLVITEGFEAREAVCSNTFLHLGGMDKVTIGDAEVHRIPIVYPLQMFPPDSRVI 181
 QY 208 PDPFNTHRSIGGFVYPPFYNTGLCHLHDCVIAPMAVALRVNVTAVAGAAHLAFD 267
 Db 182 ADFFNCNHSIGENFNYPPLFFNRPRLRLLFEAVGFAAVALRARNVDVAVAAAAHLAFD 241
 QY 268 ENHEGAVLPDDITYTYFQSSSSGTTTARGAR 298
 Db 242 ENHEGAALPADITTTAFASQG--KFORGAR 270

RESULT 6

JS-10-237-551-230

Sequence 230, Application US/10237551

Publication No. US20030165820A1

GENERAL INFORMATION:

APPLICANT: Day, Craig H.

APPLICANT: Hosken, Nancy A.

APPLICANT: Parsons, Joseph M.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND

TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION

FILE REFERENCE: 210121.538C3

CURRENT APPLICATION NUMBER: US/10/237,551

CURRENT FILING DATE: 2002-09-06

NUMBER OF SEQ ID NOS: 254

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 230

LENGTH: 248

TYPE: PRT

ORGANISM: Herpes simplex virus

JS-10-237-551-230

Query Match 11.7%; Score 736; DB 12; Length 248;
 Best Local Similarity 57.7%; Pred. No. 3.3e-64;
 Matches 139; Conservative 36; Mismatches 66; Indels 0; Gaps 0;

Y 47 PLMRNLTVKFTTSSLAUVSGARTTGLAGAGITLKLTTSHFPYSPVVFHGGKHVLPSSAA 106
 b 3 PLVGLTVESGFEANVAAVGSRITGLGTAVALSLKMPSHYSVSVVFGGRLHAPSTQA 62
 Y 107 PNLTRACNAARERFCFSRQCPVVDGAVETTGABICTFELGLEPENTILYLVVTFALFKEAV 166
 b 63 PNLTLCEARARHFGFSYARPCDLKHETTGDCALCERGLDPPRALYLVITEGFREAV 122
 Y 167 FMCNVFLHYGLDVIHNGDVIRIPLFPVQLFMPDVNRLVPDPFNTHRSIGSGFYPT 226
 b 123 CISNTFLHLGGMDKVTIGDAEVHRIPIVYPLQMFPPDSRVIADFFNCNHSIGENFNYP 182
 Y 227 PFYNTGLCHLHDCVIAPMAVALRVNVTAVAGAAHLAFDENHEGAVLPDDITYTYFQS 286
 b 193 PFFNRPLRLLEAVVGGFAAVALRARNVDVAVAAAAHLAFDENHEGAALPADITTTAFEA 242
 Y 287 S 287
 b 243 S 243

RESULT 7

S-10-156-761-9121

Sequence 9121, Application US/10156761

Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, WASHIARA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

CURRENT FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697

PRIOR FILING DATE: 2001-08-02

NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 9121

LENGTH: 850

TYPE: PRT

ORGANISM: Streptomyces avermitilis

US-10-156-761-9121

Query Match 2.0%; Score 123.5; DB 15; Length 850;
 Best Local Similarity 20.5%; Pred. No. 0.031;
 Matches 182; Conservative 104; Mismatches 327; Indels 273; Gaps 37;

QY 305 TSKPSPSGGFE-----RRLASIMAAADTALHAEVIENTGIYEBETPTDIDKEWPMFICM 355
 Db 3 TSRTTSPAGAEPPSRPPVGRRAHAQPADE-----DTGTDETVTAASADSVPVAGR 54
 QY 356 EGTLP-----LNLGSYTARVAGVIGAMVSPNSALYLTVEVSGMTE 399
 Db 55 WGVPRPTVRAKIVCLLMVPVSLALWAY-----ATVSTAQDVARLRQVQVDTTV 105
 QY 400 AKDGGPGPSFNRFYQFAGPHLAANPOTDSDGHVLSQSTGSSNTFSVDYLAL-----ICG 455
 Db 106 RAPVAAVAALQAERAAAVRHVIDFSAEPD-----SGFRTLAARTDRAVDKURLGGHTVA 161
 QY 456 FGAPLLARLLFYLERCDAGFTGGHGDALKVYTGTFDSEIPCSCKEKTRFVCAHTTVHR 515
 Db 162 DGADLPAQVGRLETFVSGA-----EQLSRLRG-----AVLER 194
 QY 516 LRQEMRFGCATQPIGVFGTWN-----SQYSDCDPLGNYPAPYLILRKPQOTENAKATMQDT 573
 Db 195 RARWDETFQYTRTTAAAFVGGALGTIGQAD-LGSDARVLL-----BFSRAGEALAQED 248
 QY 574 YRATLERLFDIEQERL-LDRGAPCSSEGLS-SVIVDHTFERRILDTLRARIEQTTQFM 631
 Db 249 AVLSSARLAGTLGGERLRLFTGAVDTRTLTDSAVD-----LSERERAAWQVATGTA 302
 QY 632 KVLVETRDYKI---REGLESEATHSMALTDP-----YSGAPCPITN 669
 Db 303 YADVRTAEKVLNAPGARITAAAPQATWPAHARVQSGMRTIEADAGRGVADRADPLTR 362
 QY 670 FLVKRTHLAVQDLALSQCHCVF-----YGOQVEGNFRNQFQVLRERFVD-----LENG 720
 Db 363 GLLTAPGAVLFGLAAVAASLIVSVRIGRGLVIELISURNALRIARKLPQAMRKLKLAG 422
 QY 721 GFISTRSITVTLSEGPVSAP-NPTLGDAPAGRTFDGDLARVSVEVIRDIRVKNRVVPSG 779
 Db 423 BEIDVR-----AEAPPGPFAEDETGQVAEALSTVHRAALRAAVE-----461
 QY 780 NCTNLSEARARLVCLASAYQOEKVDMLHGALGFLLLKQFHGLLPFGMPNPKSPNPQ 839
 Db 462 -----RAELASGISGV-----472
 QY 840 WFWTLQRNOMPADKLTHEEITIAAVKRFTEVAAINFILNPTCTIGELAQFYMANLIL 899
 Db 473 -FVNLARSQI-----LYERQLSLDSDMERSED-----PNEISDL--FLDLHLTT 515
 QY 900 KYCDHSQVLYLNTLSITIGA---RRPRDPSVLHWIRKDVTSAAD---IETQ-----945
 Db 516 RMRHRAESLI-----ILSGAAGPGRAMPVSLTNVWRAAVSEVEDYARVEVRQLPEASVV 570
 QY 946 AKALIEKTEENLPWLTTA--FTSTHLVRAAMNORP-----MVVLGISISKYHGAAGN 995
 Db 571 GAAVADLTHLMAEIVEENAAQSPPH-TRVTVGTGPNVNGYAVEVEDGLGKGTLEBAN 629
 QY 996 NRVFQAGNWSGLNGKNCVCPLETFDTRTRFIIA-----CPRGGGFCPTVGP 1041
 Db 630 RRIEQS-----EALDLFSDRLGLFWLSRAARHGKIVHLRTSPYGGTTAVVLLP 679
 QY 1042 S-----SGNRETT--LSQVIRGIIVSGGAMVOLAIYATVAVANGARA 1081

Db 680 TALLHSGTAERVPRAADTGRD-AEPAYARVAASHQSVQAVGRPA 724

RESULT 8

US-09-905-129-2
; Sequence 2, Application US/09905129
; Patent No. US20020137705A1
; GENERAL INFORMATION:
; APPLICANT: Einat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THERE
; FILE REFERENCE: 540579-2007.2
; CURRENT APPLICATION NUMBER: US/09/905,129
; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/207,821
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/084,944
; PRIOR FILING DATE: 1998-05-11
; PRIOR APPLICATION NUMBER: 60/085,673
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2

LENGTH: 2597
TYPE: PRT
ORGANISM: rattus species
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(2597)
OTHER INFORMATION: 'x' can be any amino acid

US-09-905-129-2

Query Match
Best Local Similarity 1.9%; Score 120.5; DB 10; Length 2597;
Matches 145; Conservative 113; Mismatches 287; Indels 233; Gaps 40;
2Y 416 AGPFLAANPQTD---RDGHVLSQSQTGSNTE-----FSDVYLALICG 455
Db 274 SGFLCTKTIIDPSLKSLSVTQEDNGSASTSPQDIEPFGSLNMTXXSGNKADMVCS 333
2Y 456 FGAPLLARLLFYLERCDAGATGCHGDALKYVTGTFDSEIPCCLCEKHTRPV---CAHT 511
Db 334 IQKP-----SRTSPATFEENDYIM--LNASFSTNLVCSVDYNIHQVPWQLLALYS 382
2Y 512 TVHRLRQMP-----RFGQATROPIGVGTGMSQYSDCDPLGNYPYLILRKPGD 561
Db 383 DSQLILERPQUTETPSLSRYKQVALREPDIETSIADVR-ADPFWFQOEKIVLQLNRT 441
2Y 562 QTEAAKATMQDTYRATLERLFDLEQERL-----LDR-----GAPCSSEG 601
Db 442 ATTLSTLQIOFSTDAQIALPRAEMPAERLKWMTILMNNPKLERTVLVGGTIALSCPGKG 501
2Y 602 -----LSSVIVDHPTFR-----RILDTLRARIE-QTTTFMKVL---VETRD----- 639
Db 502 DPSPHLEWLLADGSKVRAPYVSEDKRILDKNGKLEQWADSFDAGLVHCISTNDADAV 561
2Y 640 --YKIR--EGLSEATHSMALTDPPYSGAF---CPITFLVLRTHLAVQDIALSQCHCV 691
Db 562 LTYRITVVEPYGESTHDSQVHTVVTGETLDLPCLSTG-----VPDASIS---WI 608
2Y 692 FYGQVQEGNFRNQFPVLRFRFVDFLNGFGTISTRTSITVTLSEGPVS--APNPT----- 743
Db 609 LFCNTVFSQPSRDR-----QILNNGTLRILOVT-PRDQGHYQVQVAPNPGADPSS 657
2Y 744 -----LGODAPAGRTFGDLARVSEVIRDIRVRKVRVFGNCTNLSEAPAR 791
Db 658 FKVSQKKGQRMVHEDRAGSGGLGE-FNSVSLKQFASLK-----LSASALTGSEAGKQ- 711
2Y 792 LVGLASAYORQEKRVDMHLGALG-ELLKQFHGLLFRGMPNKSXSPNPFQWTLQ---R 847
Db 712 ----VSGVHRKXKHDLHRRRGDSTLRFRE--HRROLPLSARRIDPORMWALLKAKX 765

QY 848 NOMPADKLTHEBITT-----IAAVKRFTEBYAAINFILPPTCIGELAOFYMANLILK 900
Db 766 NSVP-----KKQENTTVKVPVPLAVPLVELTDEKDSAGMI--PP----- 802
QY 901 YCDHSQYLINTLTSITGARRPRDPSSVLHWIRKDVTSAAIDTQAKALLEKTENLPELM 960
Db 803 --DEEFVWLKTKASGVPRSPADSGPVNHHGFMSTIASGTEVSTVNPOTLQ-SEHLPDFK 859
QY 961 TTAFTSTHLVRAAM-----NORPMVLGISISKYHGAAGNNRVFOAGNWSG-- 1006
Db 860 LPSVINGTAVTKSMNPSIASKIEDTTONPILIP-SVAEIRDSA-----QAGRASSQS 912
QY 1007 ----LNGKRVCPLEFTRTRRPIIACPRGGFI-----CPVTG---PSSGN 1045
Db 913 AHPVTGNN---MATYGTNTYTSFTSKASTVLOPINTPESYGPOIPIITGVSRDSSD 966

RESULT 9

US-09-905-129-10
; Sequence 10, Application US/09905129
; Patent No. US20020137705A1
; GENERAL INFORMATION:
; APPLICANT: Einat, et al

; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THERE
; FILE REFERENCE: 540579-2007.2
; CURRENT APPLICATION NUMBER: US/09/905,129
; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/207,821
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/084,944
; PRIOR FILING DATE: 1998-05-11
; PRIOR APPLICATION NUMBER: 60/085,673
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 2597
; TYPE: PRT
; ORGANISM: Rattus sp.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(2597)
; OTHER INFORMATION: 'x' can be any amino acid

US-09-905-129-10

Query Match
Best Local Similarity 1.9%; Score 120.5; DB 10; Length 2597;
Matches 145; Conservative 113; Mismatches 287; Indels 233; Gaps 40;

QY 416 AGPFLAANPQTD---RDGHVLSQSQTGSNTE-----FSDVYLALICG 455
Db 274 SGFLCTKTIIDPSLKSLSVTQEDNGSASTSPQDIEPFGSLNMTXXSGNKADMVCS 333
QY 456 FGAPLLARLLFYLERCDAGATGCHGDALKYVTGTFDSEIPCCLCEKHTRPV---CAHT 511
Db 334 IQKP-----SRTSPATFEENDYIM--LNASFSTNLVCSVDYNIHQVPWQLLALYS 382
QY 512 TVHRLRQMP-----RFGQATROPIGVGTGMSQYSDCDPLGNYPYLILRKPGD 561
Db 383 DSQLILERPQUTETPSLSRYKQVALREPDIETSIADVR-ADPFWFQOEKIVLQLNRT 441
QY 562 QTEAAKATMQDTYRATLERLFDLEQERL-----LDR-----GAPCSSEG 601
Db 442 ATTLSTLQIOFSTDAQIALPRAEMPAERLKWMTILMNNPKLERTVLVGGTIALSCPGKG 501
QY 602 -----LSSVIVDHPTFR-----RILDTLRARIE-QTTTFMKVL---VETRD----- 639
Db 502 DPSPHLEWLLADGSKVRAPYVSEDKRILDKNGKLEQWADSFDAGLVHCISTNDADAV 561

LOCATION: (1)...(2597)
OTHER INFORMATION: 'x' can be any amino acid
US-09-991-630-2

Query Match 1.9%, Score 120.5; DB 10; Length 2597;
Best Local Similarity 18.6%; Pred. No. 0.46;
Matches 145; Conservative 113; Mismatches 287; Indels 233; Gaps 40;

```

QY 416 AGPHLAANPQTD--RDGHVLSQSQTGSSNTE-----FSDVYLALIG 455
DB 274 SGAFCTKPTIDPSLKSLSVTOEDNGSASTSPQDIEPFGSLNMTXXSGNKADMVCS 333
QY 456 FGAPLLARLLFYLERCDAGFTGGHDALKYVTGTFDSIPCSLCEKHTRPV-----CAHT 511
DB 334 IQKP-----SRISPTAFTENDYIM--LNASTNLVCSVDYNNHIQPVWQLLALYS 382
QY 512 TVHRLRQMP-----RFGQATRQPIGVGTWNSQYSDCDPLGNVAPYLIILKPGD 561
DB 383 DSEPLIERKQLTETPSLSRYKQVALRPEDIFTSIADV--ADPFWQOEKIVLQLNRT 441
QY 562 QTEAAKATMDTYRATLERLFDLEQERL-----LDR-----GAPCSSEG 601
DB 442 ATTLSLTIQIFSTDAQIALPRAEMRAERLKWMLMNNPKLERTVLVGGTIALSCPGKG 501
QY 602 -----LSSVIVDHPTFR-----RILDTLRARIE-OTTTOFMKVL-----VETRD----- 639
DB 502 DPSPHLEWLADGSKVRAPVYSEDGRLIDKNGKLEOMADSFAGLYHXCISTNDADADV 561
QY 640 --YKIR--EGLSEATHSMALTDPYSGAF-----CPITNVLKXTHLAVVQDLALSOCHVC 691
DB 562 LTVRIITVPEYGESTHDSGVQHVVTGETLIDPLCLSTG-----VPDASIS---WI 608
QY 692 FYCQOQVEGRNFRNQFVLRERRRVDLFGNGFISTRSITVTLSSEGPVS--APNPT----- 743
DB 609 LPGNTVFSQPSRDR-----QILNNGTILRILOVT--PKDQGHYQCVAAANPSGADFS 657
QY 744 -----LGQAPAGRTFDGDLARVSVEVIRDIRVKNRVFSGNCTNLSEAAAR 791
DB 658 FKVSQKQGVQVVEHREAGGGLGE--PNSSVSLKQPASLK-----LSASALTGSEAGKQ- 711
QY 792 LVGLASAYORQEKRVDMHLGALG--FLKQFHGLLPRGMPPNKSFPNQFWTLIQ---R 847
DB 712 ---VSGVHKXKHRLIHRRGDSTLRFRE--HRQLPLSARRIDPQWAALEKAKK 765
QY 848 NOMPADKLTHEIIT-----IAVKRTTEYAAINFNLPTTCIGELAOFYMANLILK 900
DB 766 NSVP---KKQENTTVKPVPLAVPLVELTDEKDGSMI--PP----- 802
QY 901 YCDHSQYLINTLTSITGARRPRDPSSVLHWIRKDVTSAAADIETQAKALLEKTENLPELM 960
DB 803 --DEEFWLKTKASGVGPRSPADSGPNVHGFWMTSIAGTEVSTVNPOTLQ--SEHLPDFK 859
QY 961 TTAFTSTHLVRAAM-----NORPMVVLGISISKYHGAAGNRRVFAQNWSG-- 1006
DB 860 LFSVINGTAVTKSMNPISIASKIEDITNQNPILIFF--SVAEIRDSA-----QAGRASSQS 912
QY 1007 ---LNGGKVCVCLFTDTRRRIIACPRGGFI-----CPVTG---PSSGN 1045
DB 913 AHPVTGCGN---MATYGHNTYSSFTSKASTVLQPINPTESYGVQPIPTIGVSRPSSD 966

```

RESULT 12
US-09-991-630-10
Sequence 10, Application US/09991630
Patent No. US200201515141
GENERAL INFORMATION:
APPLICANT: Elnat, et al
TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THERE
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: 540579-2007.3
CURRENT APPLICATION NUMBER: US/09/991,630
CURRENT FILING DATE: 2001-11-06
PRIOR APPLICATION NUMBER: 09/905,129

PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: 09/802,318
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 09/729,485
PRIOR FILING DATE: 2000-12-04
NUMBER OF SEQ ID NOS: 28
SOFTWARE: Patent in version 3.0
SEQ ID NO 10
LENGTH: 2597
TYPE: PRT
ORGANISM: Rattus species
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(2597)
OTHER INFORMATION: 'x' can be any amino acid
US-09-991-630-10

Query Match 1.9%, Score 120.5; DB 10; Length 2597;
Best Local Similarity 18.6%; Pred. No. 0.46;
Matches 145; Conservative 113; Mismatches 287; Indels 233; Gaps 40;

```

QY 416 AGPHLAANPQTD--RDGHVLSQSQTGSSNTE-----FSDVYLALIG 455
DB 274 SGAFCTKPTIDPSLKSLSVTOEDNGSASTSPQDIEPFGSLNMTXXSGNKADMVCS 333
QY 456 FGAPLLARLLFYLERCDAGFTGGHDALKYVTGTFDSIPCSLCEKHTRPV-----CAHT 511
DB 334 IQKP-----SRISPTAFTENDYIM--LNASTNLVCSVDYNNHIQPVWQLLALYS 382
QY 512 TVHRLRQMP-----RFGQATRQPIGVGTWNSQYSDCDPLGNVAPYLIILKPGD 561
DB 383 DSEPLIERKQLTETPSLSRYKQVALRPEDIFTSIADV--ADPFWQOEKIVLQLNRT 441
QY 562 QTEAAKATMDTYRATLERLFDLEQERL-----LDR-----GAPCSSEG 601
DB 442 ATTLSLTIQIFSTDAQIALPRAEMRAERLKWMLMNNPKLERTVLVGGTIALSCPGKG 501
QY 602 -----LSSVIVDHPTFR-----RILDTLRARIE-OTTTOFMKVL-----VETRD----- 639
DB 502 DPSPHLEWLADGSKVRAPVYSEDGRLIDKNGKLEOMADSFAGLYHXCISTNDADADV 561
QY 640 --YKIR--EGLSEATHSMALTDPYSGAF-----CPITNVLKXTHLAVVQDLALSOCHVC 691
DB 562 LTVRIITVPEYGESTHDSGVQHVVTGETLIDPLCLSTG-----VPDASIS---WI 608
QY 692 FYCQOQVEGRNFRNQFVLRERRRVDLFGNGFISTRSITVTLSSEGPVS--APNPT----- 743
DB 609 LPGNTVFSQPSRDR-----QILNNGTILRILOVT--PKDQGHYQCVAAANPSGADFS 657
QY 744 -----LGQAPAGRTFDGDLARVSVEVIRDIRVKNRVFSGNCTNLSEAAAR 791
DB 658 FKVSQKQGVQVVEHREAGGGLGE--PNSSVSLKQPASLK-----LSASALTGSEAGKQ- 711
QY 792 LVGLASAYORQEKRVDMHLGALG--FLKQFHGLLPRGMPPNKSFPNQFWTLIQ---R 847
DB 712 ---VSGVHKXKHRLIHRRGDSTLRFRE--HRQLPLSARRIDPQWAALEKAKK 765
QY 848 NOMPADKLTHEIIT-----IAVKRTTEYAAINFNLPTTCIGELAOFYMANLILK 900
DB 766 NSVP---KKQENTTVKPVPLAVPLVELTDEKDGSMI--PP----- 802
QY 901 YCDHSQYLINTLTSITGARRPRDPSSVLHWIRKDVTSAAADIETQAKALLEKTENLPELM 960
DB 803 --DEEFWLKTKASGVGPRSPADSGPNVHGFWMTSIAGTEVSTVNPOTLQ--SEHLPDFK 859
QY 961 TTAFTSTHLVRAAM-----NORPMVVLGISISKYHGAAGNRRVFAQNWSG-- 1006
DB 860 LFSVINGTAVTKSMNPISIASKIEDITNQNPILIFF--SVAEIRDSA-----QAGRASSQS 912
QY 1007 ---LNGGKVCVCLFTDTRRRIIACPRGGFI-----CPVTG---PSSGN 1045
DB 913 AHPVTGCGN---MATYGHNTYSSFTSKASTVLQPINPTESYGVQPIPTIGVSRPSSD 966

```

RESULT 13
 JS-09-991-630-13
 ; Sequence 13, Application US/09991630
 ; Patent No. US200201515141
 ; GENERAL INFORMATION:
 ; APPLICANT: Einat, et al
 ; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THERE
 ; TITLE OF INVENTION: AND USES THEREOF
 ; FILE REFERENCE: 540579-2007.3
 ; CURRENT APPLICATION NUMBER: US/09/991,630
 ; CURRENT FILING DATE: 2001-11-06
 ; PRIOR APPLICATION NUMBER: 09/905,129
 ; PRIOR FILING DATE: 2001-07-13
 ; PRIOR APPLICATION NUMBER: 09/802,318
 ; PRIOR FILING DATE: 2001-03-08
 ; PRIOR APPLICATION NUMBER: 09/729,485
 ; PRIOR FILING DATE: 2000-12-04
 ; NUMBER OF SEQ ID NOS: 28
 ; SOFTWARE: Patent in version 3.0
 ; SEQ ID NO 13
 ; LENGTH: 2597
 ; TYPE: PRT
 ; ORGANISM: Rattus species
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)..(2597)
 ; OTHER INFORMATION: 'x' can be any amino acid
 S-09-991-630-13

Query Match 1.9%; Score 120.5; DB 10; Length 2597;
 Best Local Similarity 18.6%; Pred. No. 0.46;
 Matches 145; Conservative 113; Mismatches 28; Indels 233; Gaps 40;

Y 416 AGPHLAANPQD---RDGHVLSQSTGSNTE-----FSDVYLAICG 455
 b 274 SGAFCTCTPTDPSLKSLSVTOEDNGSASTPODIEPFGSLSLNMTXXSGNADMVCS 333
 Y 456 FGAPLLARLLFYLERCDAGFTGHDGALKVVTGTFDEIPCSLCEKHTRPV---CAHT 511
 b 334 IQKP-----SRTSTATEENDIYM--LNASFNLVCSVDYVHIQVWQLLALYS 382
 Y 512 TVHRLRQMP-----RFGQATROPIGVGTMSQYSDCDPLGNAPYILRLKPGD 561
 b 383 DPLFLERKPOLTEPSSLSRYKQVALPEIDFTSIEADVR-ADPFWFQBEKIVLQNLRT 441
 Y 562 QTEAKATWQTYRATLERLFDLQERL-----LDR-----GAPCSSEG 601
 b 442 ATTLSLQIQSTDAQIALPRAEMRAERLKTMTLMNPNKLETVLVGGTIALSCPGKG 501
 Y 602 -----LSSVIVDHTFR-----RLDTLRARIE-OTTTFQMKVL---VETRD----- 639
 b 502 DPSPHLEWLLADGSKVRAPYVSEDEGRILIDKNGKLEQMAQSPDAGLYHCISTNDADV 561
 Y 640 --YKIR--EGISEATHSALTDPYSGAF-----CPITNFLVKTTHLAVQDLALSQCHCV 691
 b 562 LTYRITVVPEYGESTHDSGVQHTVVTGETLPLCLSTG-----VPDASIS---WI 608
 Y 692 FYGOVEGRNFRNFOPLVRRFRVLENGGFISTRTIVLSEGPVS--APNPT----- 743
 b 609 LFGNTVFSQPSRDR-----QILNNGTILRILOVT-PKQOHYOCVLAANFGADFS 657
 Y 744 -----LGODAPAGHTFGDLARVSEVIRDIRVKNRVVFGNCTNLSEARAR 791
 b 658 FKVSQKKGQRMVDEHREAGSGGLGE-PNSSVSLKQFASLK-----LSASALTGSEAGKQ- 711
 Y 792 LVGLASAVORCKRVMLHGLG-FLKQPHGLLFPFGMPNSKSPNPFQFWTLIQ---R 847
 b 712 ----VSGVHRKNKRDILHRRRGSTLRFRF--HRRQLPLSARRIDPQWAALEKAKK 765
 Y 848 NOMPADKLTHEEIT-----IAAVKFTTEYAAINFILPPTCIGELAQFYMANLILK 900
 b 766 NSVP-----KKQENTVTKVPVLAFLVLTDEKDSGMV--PP----- 802

QY 901 YCHSQVLYINTLTSIITGARRPRDPSSVLHWIRKQVTSAAIDETQAKALLEKTENLPELM 960
 Db 803 --DEEFWLTKKASGVPGRSPTADSGFVNEGFWTSIASGTEYSTVNPQTLO-SEHLPDFK 859
 QY 961 TTAFTSTHLVRAAM-----NORMVVLGISISKYHGAAGNRRVFOAGNWSG-- 1006
 Db 860 LFSVTNGTAVTKSMNPSIASKIEDTTNQNPFIIFP-SVAEIRDSA-----QAGRASSOS 912
 QY 1007 ---LNGGKNVCPLFTFDRTRRPIIACPRGGFI-----CPVTG---PSSGN 1045
 Db 913 AHPVTGNN-----NATYGHNTWYSSFTSKASTVLQINPTESYGQPIITGVSRPSSD 966

RESULT 14
 US-10-329-079-11
 ; Sequence 11, Application US/10329079
 ; Publication No. US20030198981A1
 ; GENERAL INFORMATION:
 ; APPLICANT: FARNET, Chris
 ; APPLICANT: ZAZOPULOS, Emmanuel
 ; APPLICANT: STAFFA, Alfredo
 ; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTIDES
 ; FILE REFERENCE: 3002-11US
 ; CURRENT APPLICATION NUMBER: US/10/329,079
 ; CURRENT FILING DATE: 2002-12-24
 ; NUMBER OF SEQ ID NOS: 66
 ; SOFTWARE: Patent in version 3.0
 ; SEQ ID NO 11
 ; LENGTH: 5245
 ; TYPE: PRT
 ; ORGANISM: Streptomyces fradiae
 US-10-329-079-11

Query Match 1.9%; Score 120; DB 12; Length 5245;
 Best Local Similarity 20.2%; Pred. No. 1.8;
 Matches 262; Conservative 138; Mismatches 434; Indels 464; Gaps 68;

QY 102 PSSAAPNLTRACNAAREPFGSRGCGQP-----VDGAVTTGAEICTRL-----GL 147
 Db 464 PAAEPADEGLEAVCDTFARQAATPEAPVGGPVALTFAEADARVSRRLRLISGA 523
 QY 148 EPENTILVTVLTALFEKAVFMCNVPFLHYGLDVIHNGDVIRIPLFPVQLFMDVNRV 207
 Db 524 GPE---VRVAVCLDNALMPTTV-----LAVLSGAVHVP----- 556
 QY 208 PDPFNTHRSIGEGFVPTFPFNTGLCHLIHDCVIAPMAVALRVNRVTVARGAHLAF- 266
 Db 557 -DPRSPHER-----LAAVERD--VAPLLVLAERATEAAVADLAAPVLVL 597
 QY 267 -DENHEGAV-----LPPDITYVFSQSSSG-----TTTARG-----A 297
 Db 598 DDPSTEALDLDPOVTDARTAPLPGHAAYVHTSGSTGRPKGVTVDRHGRLLQA 657
 QY 298 RNDVNSTSKPSPG-GFERLASI-----MAADTALHA----- 330
 Db 658 HRVTFSRIRPSAGGPGRAAAVVSFSDASDPLLANVAGHELAWIDELRFPDPPVAY 717
 QY 331 -----EVIPTGYEE--TPT-----DIKEWPMFIGMGTLPRLNAL 365
 Db 718 FDRRIDVYDUTPTVFRSLDAGLLEEGFPCPSVALGGEAMDGELWERLRAAAPRYTAM 777
 QY 366 GSY-----TARVA-----GVIGAMVFPNSALVLTVEDESGMTEAKDGGPSPFN 411
 Db 778 NYTGTETAVDAVTVGLDPLPGTIGRPV--PRWAY--VLDAGLPVPGVLGELY-- 830
 QY 412 FYQFAGPHLANPOTDRDGHVLSQSTGSSNTSPDYALICGFGAP-----LLAR 463
 Db 831 ---LAGPVGARGYLQ---HALTAER-----FVACPFGPGERMVRTGLAR 871
 QY 464 L-----LFLVLRCD-----AGAF-----TGGHGDALKYV 487
 Db 872 WLPDGLHYVVGDEBQVKIRFRIEPOGEVAALRELEGVAAAATVREDTPTGTRRLVGYV 931

```

; SEQ ID NO 8
; LENGTH: 19608
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.
; US-10-084-846A-8

Query Match      1.9%; Score 118; DB 12; Length 19608;
Best Local Similarity 19.7%; Pred. No. 30;
Matches 240; Conservative 133; Mismatches 443; Indels 404; Gaps 62;

QY 106 APLTFCACVAARERGFSEC---QGPVVDGAVETTGAE---ICTRLGLEP----- 149
DB 11736 APLRSTRTGAPRGHARRCRGRGPIGA---RGAEPHSHSRRLRRSPDSAPSLL 11792
QY 150 ---ENTILYLVVTFALFKEAVFMCNVLFLHYGGDLIVH---INHGDI-----VIRIPLPP 195
DB 11793 ENLENNAKQIRISAKGPA-----GGLSPSHGCCRHQSHAVGQALVQRLPFVP 11842
QY 196 VQLFMPD-----VNRLVDPDPNTHRRSIGGFVY-PTPFYNT 231
DB 11843 V-LGVPAHGVGEAL-ERHTGCVARAEAGVAORVAVV-----SLPVGITLDPVPAQA 11895
QY 232 GLCHLHDCVIAPMAVALRVNVTVAVGAHAFDENHE-----GAVLP-PDITTYTFQ 285
DB 11896 GVQQLLRQIIGOLG---RARDVVDLPR---QPAFGDQDPAPHVGVHPVDPVAAIPVQ 11949
QY 286 SSSGTTTARGARRNDVNSTKPSGSGFERLAS---INAAATLHAEVIFNTGIVEETP 343
DB 11950 RDAVAVQPGGGEORNDL-----LRELIGVVVVAAPRAHRQ----- 11985
QY 344 TDIKEMPMFTGMEGTLPRLNA-LGSYTARVAG-----VIGAMVFPNSA 386
DB 11986 ---PMGAGV-GEQDHVAARLGRVGVGDRRVLPGARLDAVAHLVGGDLHDPGRA 12038
QY 387 ---LYL-TEVEDS---GMTRAKOGGPGSPNRFVQ-----PAGPHLAANPQTRDRGHVLS 434
DB 12039 GLQAHLECLDALHIGTYERRRVGDRFVDFRGREVDHQIVAGDHLGEOFGI-ADVAVHE 12097
QY 435 SQSTGSSNTSFSDYLALICGFCAPILLARLLFYLERCDAGFTGGHGDALKVYTG----- 489
DB 12098 GVPAGDRLQVRQVPGVQQLVQHCHPRIAV-----PDQFAHVLGADEAR 12141
QY 490 -TFDSEIPCSLCEKHTRPVCAHTTVHRLRQMRPFQOATROPFGVGTWNSQVSDCDPLG 548
DB 12142 STDDQDVPSFP---RTRAAVSHVPSKLRHRRVRLRRPVQPTWT-RLWSPSTSRCAVDV 12197
QY 549 NYAPYLIRKPGDQTEAAKATMDTYRATLERLFIDLEQERLLD---RGAPCSSEGLSS 604
DB 12198 RGAAPTSSVRPIETESSTRATPSIEAPRSTIE-YWISLCATROLPSMEVKGPRESCVTVSQ 12256
QY 605 VIVDHPTRFR-----ILDTLR--ARIEQTQTFMKVLVETRDYKIREGLSEATHMA-- 654
DB 12257 PMTAGTPWRPELIFAPASTTTRTSSLASSTSPSMRV---SRDSSRCRTSRTSVTLATS 12313
QY 655 ---LTFDPYSGAFCDITNPLVKRTHLAVVQDLALSOQCHCYFGQVQGVGRNFRNQ----- 705
DB 12314 PQYRSITVEPTS---CP-----WSASHC---RASVITSSFRQGRMP 12349
QY 706 ---FQPLRRRFVDLPNGGFISTRITVLTSEGPVSAPNPTL-GODAPAGKT 753
DB 12350 ATACHTAELKAYETIARSL--LGSWGF-STTAVTRPFS--SSATPNLRGSATRASMT 12403
QY 754 FDGLAR---VSVSEVIRDIRVQRVFPVSGNCTNLSEARLVLGLASAYQQRKRYDMLH 810
DB 12404 CAAGLSRRKSAKSTKMPSTLSPRYM-----TKSSSPRKSDTSTASAPR----- 12449
QY 811 GAGFLKQFHLGLFPGRGMPNSKSNPQFW-----TLQONQMPADKL---THEEIT 861
DB 12450 GASCMWMTS-----RPNB-APSTEAWSAAVAPTITSDIPASIMESSTERTV 12498

```

```

; RESULT 15
; US-10-084-846A-8
; Sequence 8, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MOHLENWEG, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084, 846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent In Ver. 3.2

```

us-09-769-699-2.rapb

Fri Jan 30 16:10:01 2004

862 TIAAVKPTTEYAAI--NFINLPPTCIGELAQFYMANLILKYCDHSQVLYINTLTSITGA 919
12499 LLATGTSCLAEVCVIGRSLEPTPT-----NMSAFISHTLSA--- 12536
920 RPRDPSSVLHWIRKDVTSAAIDETQAKALLEKTENLPELWTTAFTSTHLYVRAAMQRP 979
12537 -----PSDQV--VSRELQMASNVSAQORSYWRVQVFFRPGGLARPPHPAYPCHELSQ 12589
980 VV-----LGISISKYHGAAGNNRVFOAGNWSGLNGKNCVPLFTFDSTRFIIACPRG 1032
12590 CVQDAWPSAVGCCSVRR---SVRTNLSLWVGSFQKMLAGTS-----RPRF----- 12631
1033 GFICPVTGSSGNRETTLSQDVRGIIIVSGAMVQLAIYATVVRVAVGARACHMAFDWL 1092
12632 -----SSRRRTTIVITESPCSTSG-----RRTCAVAGASRISCAI---CSR 12670
1093 TDDEFLARDLEELHDQIIQTLETPWVEGLEAVKILDEKTTAGDGETTNLAFNFDSC 1152
12671 TSSS-----TRESSSGD-----R 12684
1153 PSDDTTSNVLNIGSNISGS 1172
12685 PANDTTSASASPPGSSVSGA 12704

Search completed: January 30, 2004, 13:20:11
Job time : 49 secs

1903590KP1



ion date: 03-03-2004
 king Officer: HABDULEKADER - HUDA ABDULEKADER
 n: OIPEScanning
 sier: 10190359

al Date: 09-16-2002

o.	Doccode	Number of pages
	NPL	12 /
	NPL	4 /
	NPL	5 /
	NPL	8 /
	NPL	9 /
	NPL	4 /
	NPL	5 /
	NPL	2 /
	NPL	13 /
	NPL	6 /
	NPL	14 /
	NPL	6 /
	NPL	8 /
	NPL	40 /
	NPL	6 /
	NPL	7 /
	NPL	8 /
	NPL	4 /
	NPL	3 /
	NPL	4 /
	NPL	30 /
	NPL	8 /
	NPL	4 /
	NPL	8 /
	NPL	17 /
	NPL	14 /
	NPL	5 /
	NPL	9 /
	NPL	12 /
	NPL	7 /
	NPL	10 /
	NPL	
	NPL	
	NPL	
	NPL	